Allosteric regulation of 3 deoxy-D-*arabino*-heptulosonate 7-phosphate synthase

---

A thesis submitted in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy in Biochemistry

at the University of Canterbury

by

Yifei Fan

2017

Abstract

Allostery refers to the process in which interaction of an effector ligand with one site of the protein changes the function of the protein at a distant site. Despite the critical role of allostery in regulation of metabolic pathways, little is known about the details of the allosteric networks and the remarkable diversity in allosteric mechanisms. This study utilises several examples of allosteric proteins to illustrate the interwoven relationships between various allosteric mechanisms, ranging from large conformational changes to subtle dynamic communications. An important metabolic enzyme, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAH7PS), was selected for this study due to its unique diversity of allosteric regulations and its important role as the first committed step in aromatic amino acid biosynthesis. Essential in many pathogenic bacteria, however lacking mammalian counterparts, DAH7PS and related pathway enzymes provide opportunities in development of novel antimicrobial drugs.

The first part of this study addresses the determinants of allosteric ligand selectivity and potency for DAH7PS enzymes that exhibit large conformational changes, and provides structural and functional insights that contribute to the understanding of the role of conformational change in allostery. The second part of this study addresses interchangeability between two different allosteric mechanisms by demonstrating the ease of gene fusion to link two contemporary protein domains and produce functional chimera. The third part of this study addresses the allosteric regulation in DAH7PS enzymes from a different subfamily, for which no large conformational changes are involved in delivering allosteric communication. The crystal structure of a related chorismate mutase enzyme contributes to the understanding of protein-protein interactions associated with allosteric regulations employed by this type of

DAH7PS. The final part of this study addresses the current limitations in studying allosteric systems and explores the advantages of new techniques, including Förster resonance energy transfer and electron paramagnetic resonance, in offering valuable information on the timescales and molecular structures associated with allostery.

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Deputy Vice-Chancellor's Office
Postgraduate Office

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 3

Fan Y, Cross PJ, Jameson GB, Parker EJ: **Interchangeable regulatory domains: exploring modular allostery en route to chorismate**. Submitted to *Proceedings of the National Academy of Sciences* 2017.

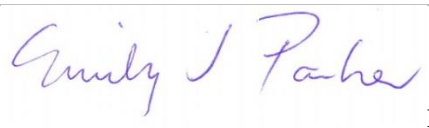
Please detail the nature and extent (%) of contribution by the candidate:

80%

The candidate designed and performed the research. The manuscript was drafted by the candidate, and edited and contributed to by co-authors.

Certification by Co-authors:
If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text



Name: *Emily Parker* Signature: Date: *11 Oct 2017*

Table of contents

List of figures

List of tables

Abbreviations

ACT	aspartate kinase, chorismate mutase and TyrA
Ape	*Aeropyrum pernix*
AUC	analytical ultracentrifugation
BSA	bovine serum albumin
Bsu	*Bacillus subtilis*
BTP	1,3-bis[tris(hydroxymethyl)methylamino]propane
Cd	cadaverine
CD	circular dichroism
CM	chorismate mutase
DAH7PS	3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase
DEER	double electron–electron resonance
DNA	deoxyribonucleic acid
DSF	differential scanning fluorimetry
DTT	dithiothreitol
E4P	erythrose 4-phosphate
Eco	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EPR	electron paramagnetic resonance
FRET	Förster resonance energy transfer

Gsp	*Geobacillus sp.*
GST	glutathione sulfur-transferase
HMP	hydroxymethyl pyrimidine
Hpy	*Helicobacter pylori*
IMAC	immobilised metal affinity chromatography
IPL	isochorismate-pyruvate lyase
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
KDO8PS	3-deoxy-D-*manno*-octulosonate 8-phosphate synthase
Lmo	*Listeria monocytogenes*
MS	mass spectrometry
mTG	microbial transglutaminase
Mtu	*Mycobacterium tuberculosis*
Nme	*Neisseria meningitidis*
NMR	nuclear magnetic resonance
Pae	*Pseudomonas aeruginosa*
PCR	polymerase chain reaction
PDB	Protein Data Bank
PELDOR	pulsed electron–electron double resonance
PEP	phosphoenolpyruvate
Pfu	*Pyrococcus furiosus*

PHB	*para*-hydroxybenzoate
PRE	prephenate
PYR	pyruvate
RMSD	root mean square deviation
SAXS	small-angle X-ray scattering
SCA	statistical coupling analysis
SD	standard deviation
SDM	site-directed mutagenesis
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	sedimentation equilibrium
SEC	size exclusion chromatography
SOC	super optimal broth
SV	sedimentation velocity
TAE	tris-acetate-EDTA
TCEP	tris(2-carboxyethyl)phosphine
TE	tris-EDTA
TEV	tobacco etch virus
TIM	triosephosphate isomerase
Tma	*Thermotoga maritima*
Tris	tris(hydroxymethyl)aminomethane
Tye	*Thermodesulfovibrio yellowstonii*

Acknowledgements

I would like to express my gratitude to a number of people who have contributed to the completion of this thesis. First and foremost, I am greatly indebted to my supervisor Emily Parker for her support, guidance and encouragement I have received over the years. She has been a true inspiration for me on both professional and personal levels since we first met over seven years ago. I have learnt so much from her enthusiasm, knowledge, and experience. I am incredibly fortunate to have her guide my way through the early stages of my scientific career. Thank you for everything, Emily.

It has been a privilege working alongside an excellent group of scientists of the Parker lab, all of whom have contributed to this thesis in one way or another. Thanks to Dr Ali Nazmi and Dr Gerd Mittelstädt for helping me to get started in the lab and providing mentorship in molecular biology and protein works. Thanks to Dr Wanting Jiao for her contribution of computational studies on various projects. Her helpful feedback, good chats and wine were always appreciated. Thanks to Dr Eric Lang for his involvement in multiple projects and our pleasant discussions on everything. Thanks to Dr Leyla Bustamante for making the lab a better place to work and for her warm hugs. Thanks to Fiona Given and Yu Bai for our stimulating discussions, and Kyle Van de Bittner, Nicola Blackmore, Parastoo Khajeaian for their friendship and company through ups and downs of this journey.

I have benefitted from interactions with a number of people outside the Parker lab. I would like to thank Dr Andy Pratt, who provided helpful advice throughout my time at UC and encouraged

me to apply for a PhD position. I am grateful for his time and kind help on course selections during my studies overseas, his wisdom and helpful discussions around my project. Professor Geoffrey Jameson provided valuable knowledge around protein crystallography and contributed to various manuscripts. I deeply appreciate his constructive feedback, input and interesting discussions. Thank you to Dr Matt Polson, for his help with various instruments, continued patience and always friendly attitude. Thank you to Professor Sally Gaw for her continuous encouragement over the years. I would also like to thank Dr Amelia Albrett and Wayne Mackay for providing helpful technical support.

Thank you to my collaborators Professor Jeffrey Keillor, Dr Alistair Fielding and their research groups, for their efficient communications, advice, and work on FRET and EPR related experiments.

There are several organisations and people associated with which that I would like to acknowledge. I would like to thank UC and the Department of Chemistry for providing me with a doctoral scholarship and various financial support over the years, NZ Synchrotron Group for providing wonderful training and learning opportunities in Japan and Australia, Maurice Wilkins Centre for conference opportunities, and Biomolecular Interaction Centre for providing equipment and training. I also truly appreciate the help and friendship I received from all the past and present staff at the Australian Synchrotron MX and SAXS beamlines, particularly Dr Tim Ryan for his knowledge, patience and enjoyable discussions.

Finally, thanks to my family, especially my mother, for her support in everything I do. A special thank you to Andrew, for your understanding, your patience, your love and encouragement, and for always being there when things got tough. I could not have done this without you.

Chapter 1. Introduction and overview

1.1. Introduction

Control of biosynthesis and degradation of metabolites is an essential part of cellular function to ensure normal growth and maintenance of the cell by properly regulating the level of these compounds to satisfy various metabolic demands. This fundamental biological process often involves regulation of important proteins via intricate networks of interacting molecules.[1] Allosteric regulation is one of the most common mechanisms to achieve direct control of protein function, where interaction of an effector ligand molecule with the target protein introduces changes in protein conformation and dynamics to modulate the interaction of another molecule at a distant functional site.

1.1.1. Evolving concept of allostery

Allostery as one of the most fundamental and intriguing biological processes in understanding cellular chemistry, signalling, metabolism and diseases has drawn increasing attention in the central focus of biological research; hence the concept of allostery and its mechanisms continue to evolve since it was first articulated more than half century ago.[2-3] Some key experiments leading to an understanding of allostery were documented in the early 1900s, where Bohr effect and Hill equation were described to report the sigmoidal binding affinity of haemoglobin to O_2. In 1958, the birth of the first X-ray structure of sperm whale myoglobin revolutionised the understanding of protein structure.[4] Soon after, the discovery of feedback inhibition of enzymes

and naming of regulatory sites as ‘allosteric sites’ marked the beginning of allostery.[5-6] In 1965 and 1966, the two classical models, the Monod-Wyman-Changeux (MWC) model and the Koshland-Nemethy-Filmer (KNF) model, were introduced to describe ligand binding at one site causing conformational change of protein that alters its function at the second site.[2-3] The MWC model focused on the transition between two pre-existing, distinct conformational states (relaxed-R and tensed-T),[2] whereas the KNF model emphasised on the sequential change and ‘induced-fit’ from ligand binding at the first site.[3] Both of these rested on the assumption that only two defined conformations can exist, and in the presence of ligand, major changes occur in the shape of the protein.

Although the concept of the two models has been proven to be successful at describing basic features of the conformational transition,[7-9] numerous theoretical realisations and experimental observations have demanded the development of more detailed mechanisms (how) and energetics (why) in understanding allosteric communication. In 1984, the idea that allostery can occur in the absence of conformational change was first proposed based on a theoretical proposition.[10] Cooper and Dryden demonstrated that changes in amplitude and frequency of protein thermal motions can occur upon allosteric interaction, without changing the average atomic positions of the protein structure. This important update on the definition of allostery in thermodynamic terms laid the foundation for modern concept of allostery and aided in understanding of allosteric mechanisms.

1.1.2. Allostery is structured yet dynamic

The origin of allostery was based on the classic perspective that an allosteric protein is structured and the propagation of allosteric signal follows a defined pathway. The current understanding of allostery has been significantly advanced with additional structural and dynamics data provided by the development of new techniques, such as nuclear magnetic resonance (NMR) spectroscopy and Förster resonance energy transfer (FRET), supplemented by theoretical analyses and simulations. Copper and Dryden's prediction of allostery without conformational change was validated 20 years later by direct NMR data proving that allostery can be mediated purely by transmitted changes in protein dynamics without any conformational change.[11] This important experiment used relaxation dispersion NMR and NMR-detected hydrogen exchange to probe the allosteric details of the dimeric catabolite activator protein (CAP) in response to its ligand cyclic AMP (cAMP) (Figure 1.1). Binding of cAMP to one CAP chain was shown to have no conformational effect on the other chain. However, the first cAMP partially enhanced the dynamics of the protein, whereas the second cAMP completely quenched system dynamics. An entropy penalty for the second cAMP was therefore arose from quenching of protein dynamics upon ligand binding leading to the negative cooperative binding of cAMP to CAP. Several studies have demonstrated the change in conformational entropy upon cAMP binding via backbone and side-chain dynamics.[11-13]

The theory of dynamic-driven allostery was verified by another important study where a distal element of PDZ domain was shown to significantly affect the side-chain dynamics of the ligand binding site without global changes in the backbone as revealed by NMR experiments. This effect was almost solely entropic in nature as determined by isothermal titration calorimetry (ITC).[14] Although it was generally accepted that changes in enthalpy (contributed by large conformational change) of some protein systems were essential to transmit allosteric signals,

these studies on the CAP and PDZ systems revealed the important role of dynamics (entropy) in allosteric mechanism.

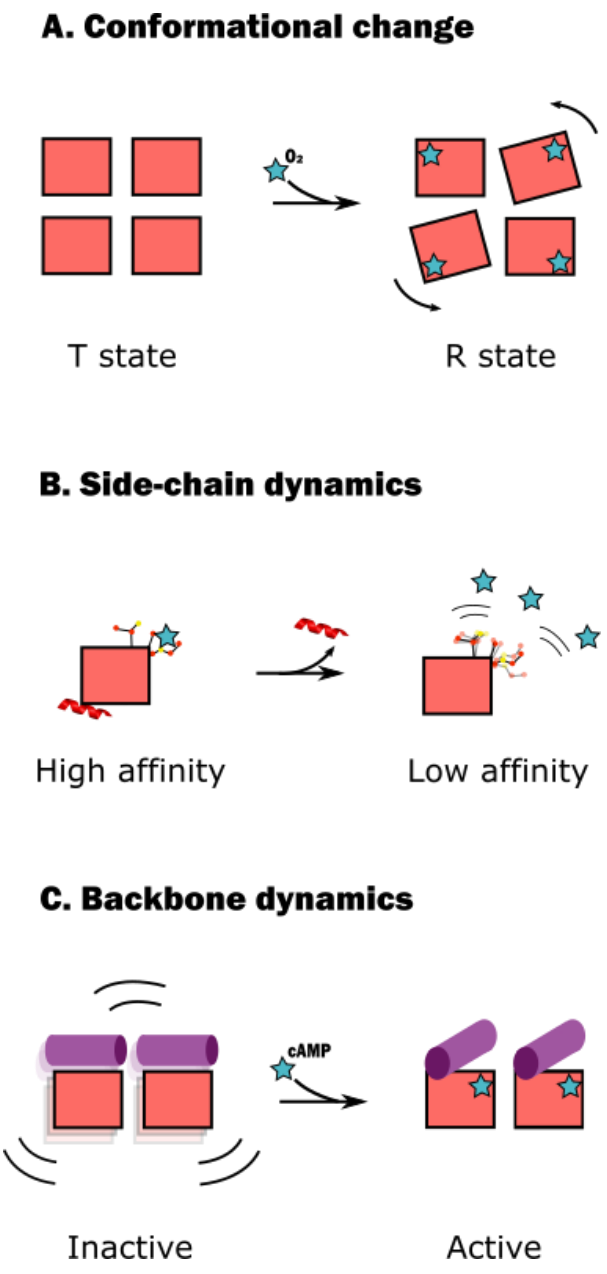


Figure 1.1. Structured and dynamic allostery. A. Binding of oxygen to T state of haemoglobin causes conformational shift of the subunits and drives the equilibrium towards R state.[15-16] B. A remote α helix of PDZ domain significantly affects the side-chain dynamics at the ligand binding site. Truncation of this helix caused dramatically increased dynamics and reduced ligand affinity.[14] C. Binding of ligand

cAMP to CAP quenches backbone dynamics and allows appropriate rearrangement of the functional domain to activate transcription.[11-13]

1.2. Aromatic amino acids

The three aromatic amino acids phenylalanine (Phe), tryptophan (Trp), and tyrosine (Tyr) are essential aromatic compounds in metabolism. They are required for protein biosynthesis, constituting almost 10% of proteins on average, and play critical roles as precursors for many important metabolites involved in cellular growth, regulation and maintenance.[17]

1.2.1. Biosynthesis of aromatic amino acids

Phe is a common precursor of many phenolic compounds, such as lignin and flavonoids.[18] Phe-derived compounds constitute almost 30% of organic matter in some plants, hence majority of the carbon flux is often directed to Phe.[19-20] Trp is a precursor of phytoalexins, alkaloids, and some plant hormones. Tyr is important in synthesis of pigment betalains and some quinones.[21-22] All three aromatic amino acids are derived from the shikimate pathway in most microorganisms, fungi and plants (Figure 1.2). However, this pathway is never identified in animals. Phe and Trp are essential amino acids for humans and livestock, hence must be supplied through diet. Tyr can be synthesised in mammals, however a source of Phe is required for an enzymatic hydroxylation step.[23-24] Trp and Tyr are also precursors of serotonin and some neurotransmitters in animals.[25]

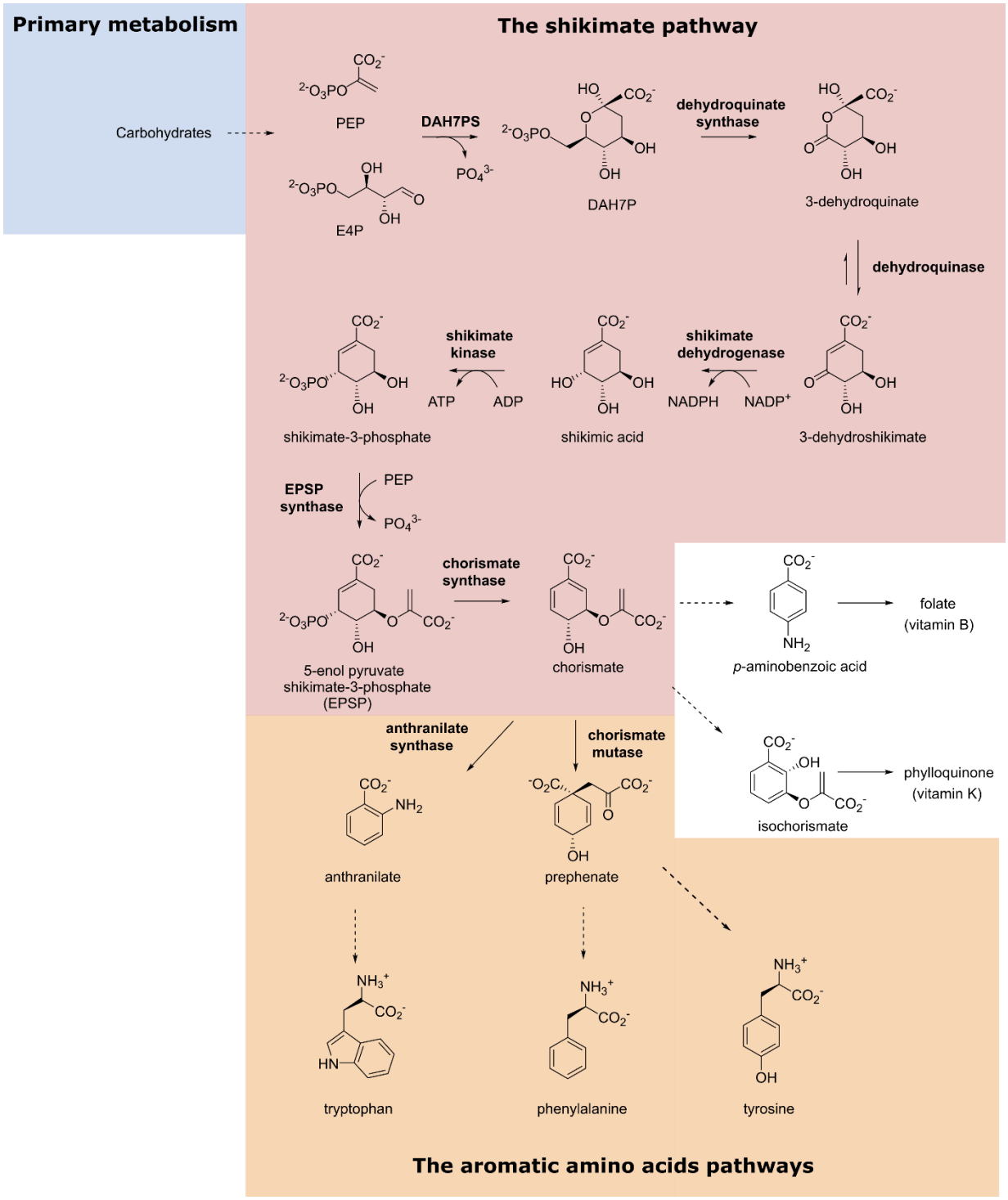


Figure 1.2. The shikimate pathway linking metabolism of carbohydrates to biosynthesis of aromatic compounds. The pathway produces final product chorismate, a common precursor for Trp, Phe and Tyr, as well as other important metabolites. Reactions that involve multiple steps are shown as dashed arrows. DAH7PS, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase.

The shikimate pathway links central carbon metabolism and the aromatic amino acid networks by converting two carbohydrate precursors phosphoenolpyruvate (PEP) and erythrose 4-phosphate (E4P) to ultimately produce the end-product of the pathway, chorismate, via seven enzyme-catalysed reactions (Figure 1.2). From the universal precursor chorismate, the pathway branches towards the formation of prephenate and anthranilate leading to aromatic amino acids and other important compounds such as vitamins K_1 and B_9.[23-24, 26-27] Furthermore, other intermediates of the pathway also serve as starting points for biosynthesis of numerous secondary products. Clearly, the shikimate pathway is of substantial importance to the biosynthesis and regulation of many compounds of commercial and medical interest.

1.2.2. Regulation of the shikimate pathway

Regulation of key metabolites by enzymatic pathways is vital for maintaining normal metabolism of the cell. Regulation of the shikimate pathway is particularly crucial as the synthesis of aromatic amino acids is energetically expensive. The three aromatic amino acids represent almost the entire output of aromatic biosynthesis in most prokaryotes.[26] Activity of the shikimate pathway is often monitored by intracellular levels of Phe, Trp, Tyr and other pathway intermediates.[26]

Metabolic networks are often constituted by branched reaction pathways, for which the regulation usually occurs at the committed step of the pathway by interwoven regulatory mechanisms controlling gene expression and/or enzyme activity. Specifically, the enzyme 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAH7PS), which catalyses the first step of the shikimate pathway, is frequently controlled at both transcriptional level and protein level by feedback regulation.[23] Feedback regulation is found consistently throughout amino

acid metabolism, in which pathway end-products or intermediates inhibit or activate the activity of the early enzymes in the pathway to reduce the waste of metabolites, deliver tight control of biosynthesis, and ultimately increase the fitness of the cell.[28]

Furthermore, regulation of enzymes located at branch points of the aromatic amino acid pathway is important for fine-tuning the distribution of metabolites in response to cellular requirements. For instance, anthranilate synthase, which converts the branch-point intermediate, chorismate, into anthranilate en route to biosynthesis of Trp, is regulated at transcriptional level and allosterically regulated by Trp.[29] Chorismate mutase (CM), which converts chorismate into prephenate, is also feedback regulated by Phe and/or Tyr in some organisms.[30-31] Besides the primary regulation at the committed step and at branch-points, regulation of other enzymes of the pathway is also noted, such as the gene regulation of 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase at the sixth step of the pathway.[23]

1.2.3. Applications of the shikimate pathway

The biosynthetic enzymes comprised in the shikimate pathway are shown to be essential in many organisms. Due to the absence of the pathway in mammals, all of the enzymatic steps offer important targets for the development of live vaccines, antibiotics, and herbicides.[32-34] For example, mutants of several enzymes in the pathway were reported to attenuate the growth of *Salmonella enterica* and eliminate its infection in mice.[35-38] In addition, knockout studies identified shikimate enzymes that were essential to the virulence of *Shigella flexneri.*[39-40] Similar studies on modifications of various shikimate enzymes in a number of pathogenic bacteria have shown impaired pathogen growth or nonvirulence in mice.[41-45]

The importance of the shikimate pathway is also demonstrated by the development of enzyme inhibitors targeting pathway enzymes. One of the early examples was discovered in the herbicide field by Jaworski's group in 1970s.[46] Their lead compound, glyphosate, which targets EPSP synthase, is a dominantly used herbicide worldwide. This billion-dollar success is due to its highly effective broad-spectrum in weed management, and at the same time it is environmentally and toxicologically safe.[47] In fact, glyphosate in combination with pyrimethamine has also been tested effectively in mice as an anti-parasite agent to limit pathogenic infections including toxoplasmosis and malaria.[48]

The shikimate pathway has also attracted growing research interest in the field of bioengineering and biotechnology, as the intermediates and derivatives of the pathway can be utilised for biosynthesis of numerous compounds with a wide range of chemical properties or biological activities. For example, extensive research efforts have been made for development of highly efficient methods for production of aromatic amino acids in *Escherichia coli* due to their frequent use in dietary supplements as well as industrial products such as artificial sweetener aspartame derived from Phe.[49-51] The shikimate pathway has also been artificially engineered to supply target molecules or precursors for biosynthesis of compounds with pharmaceutical activities. Salicylic acid as one of these compounds is the active ingredient of aspirin, which is regularly used to treat pain, fever, and inflammation.[52] Its production can be achieved at gram per litre scale via an engineered shikimate pathway in *E. coli*, where metabolic influx is improved by overexpression of a number of enzymes in the pathway.[53-54] The production of compounds with anticancer activities, including alkaloids,[55] flavonoids,[56-57] violacein, and deoxyviolacein,[58-59] has also been achieved and improved by rational design of the shikimate pathway.

1.3. DAH7PS

The enzyme 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAH7PS) is the first enzyme of the shikimate pathway, catalysing the divalent-metal dependent condensation of PEP and E4P to produce DAH7P and inorganic phosphate. DAH7PS is essential for most microorganisms due to its key role in the synthesis of chorismate. This enzyme, situated at the committed step of the pathway, is precisely feedback regulated in many organisms because of the necessity to efficiently control the pathway flux in response to changes in the level of pathway outputs.[60-64] Due to this critical role, a deep understanding of the regulatory mechanisms utilised by DAH7PS is of primary importance for development of pharmaceutical and industrial products.

1.3.1. Classifications and structures of DAH7PS

DAH7PS enzymes are classified into two distinct types: Type I and Type II, on the basis of their distinctive amino acid sequence and protein size.[65] Type I shares less than 10% sequence identity with its Type II counterparts. Type I DAH7PSs are generally smaller than Type II enzymes, with molecular weights less than 40 kDa, whereas Type II enzymes are generally larger than 50 kDa. Type I enzymes are further divided into two subfamilies based on sequence similarity: Iα and Iβ (Figure 1.3). Type Iβ DAH7PSs share low sequence identity of less than 18% with Type Iα, but are more closely related to 3-deoxy-D-*manno*-octulosonate-8-phosphate synthase (KDO8PS). KDO8PS catalyses a similar aldol-like condensation reaction between PEP and arabinose 5-phosphate (A5P) to yield KDO8P, and is involved in biosynthesis of the cell wall in Gram-negative bacteria.[66-67]

Since the first crystal structure of DAH7PS was reported in 1999, there have been more than 60 DAH7PS structures deposited to Protein Data Bank (PDB) to date. Crystal structures of DAH7PS have been solved for members of all subtypes, including Type Iα DAH7PSs from *E. coli*,[60, 68] *Saccharomyces cerevisiae*,[62] *Francisella tularensis*,[69] and *Neisseria meningitidis*;[70] Type Iβ enzymes from *Pyroccocus furiosus* (*Pfu*),[71, 72] *Aeropyrum pernix* (*Ape*),[73] *Bacillus subtilis* (*Bsu*),[74-75] *Thermotoga maritima* (*Tma*),[63, 76] *Listeria monocytogenes* (*Lmo*),[77] and *Geobacillus sp.* (*Gsp*);[78] as well as Type II enzymes from *Mycobacterium tuberculosis* (*Mtu*)[64] and *Corynebacterium glutamicum*.[79]

Although sequence similarity is low between Type I and Type II enzymes, the catalytic domain of all available structures adopts a $(\beta/\alpha)_8$ barrel core housing the active site situated towards the C-terminus of the barrel. Variations amongst different subtypes of DAH7PS arise from structural extensions or additions to the core barrel. Type Iα typically contains a double β-sheet insertion and a double α-helix extension at the N-terminus. Type Iβ DAH7PS comprises solely the core barrel as observed in *Pfu*DAH7PS and *Ape*DAH7PS,[71, 73] or two domains with the catalytic barrel attached to either a chorismate mutase (CM) domain as seen in *Gsp*DAH7PS,[78] or an ACT domain as reported in *Tma*DAH7PS.[63] Type II enzymes display extensive modifications to the core structure generally with a double-α-helix insertion and a large N-terminal extension.[64] These structural decorations to the DAH7PS catalytic core are believed to determine the allosteric regulation of the enzymatic activity.[72, 80]

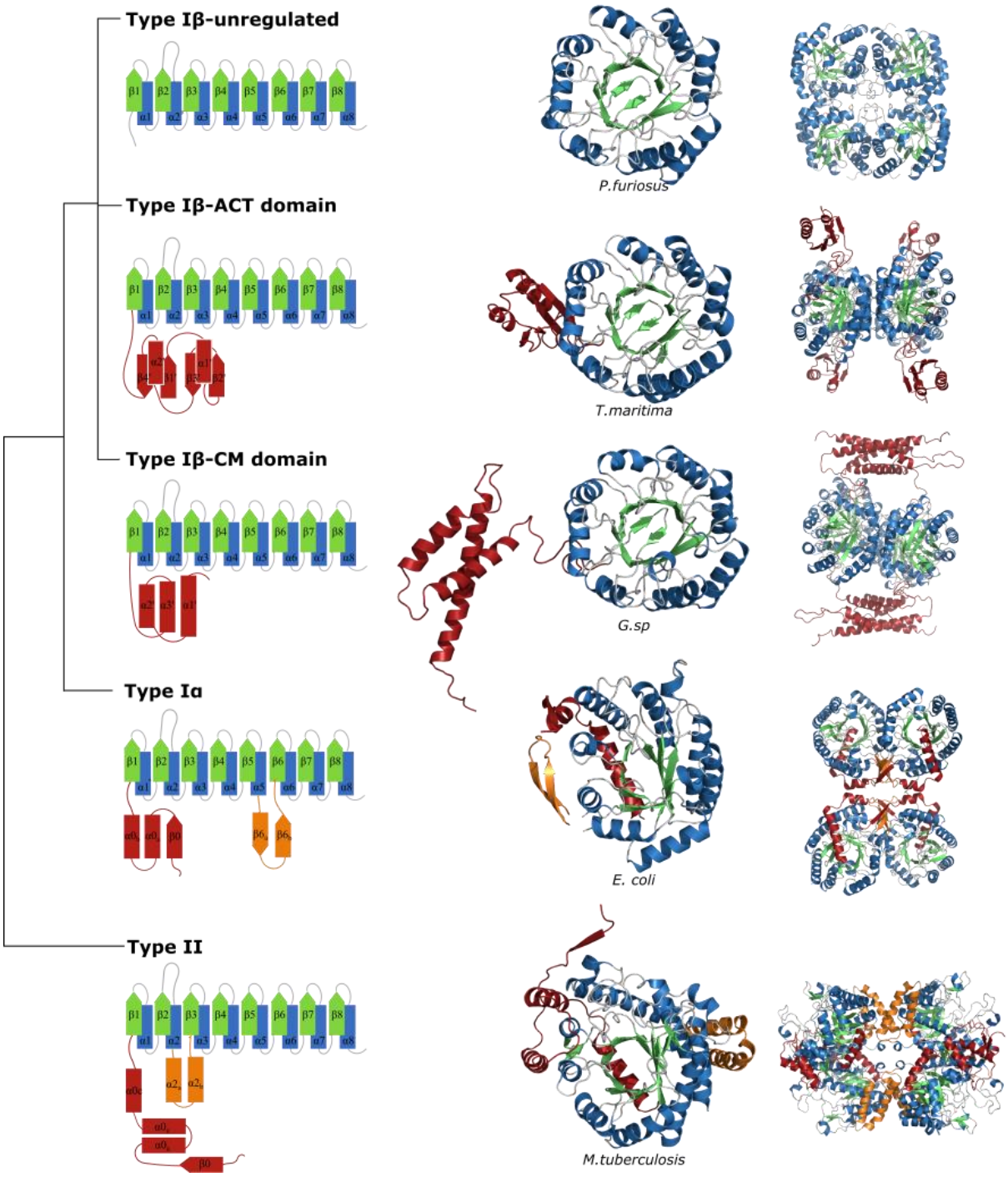


Figure 1.3. Classification of DAH7PS. DAH7PS is classified into Type I and Type II. Type I is further divided into Type Iα and Iβ. The simplest DAH7PS, comprised of the catalytic barrel only, is unregulated. Monomeric unit and tetrameric assembly are shown for each protein. The α-helices in the core barrel are shown in blue, β-sheets are green. N-terminal extensions are red, and insertions are orange.

1.3.2. Allosteric regulation of DAH7PS

An array of mechanisms for modulating DAH7PS activity in response to downstream pathway product concentrations have been delivered and maintained through the course of evolution. Numerous studies have identified the intricate networks of interwoven transcriptional and allosteric regulations that interact to control DAH7PS activity.[81-83] Although the variability in transcriptional control amongst different organisms is intriguing, the remarkable diversity of the allosteric mechanisms employed by DAH7PS enzymes is truly exceptional and perhaps the most researched for any enzyme.[24] Ranging from large conformational changes to delicate changes in protein dynamics associated with regulation, DAH7PS is an excellent model system for in-depth study of enzyme allostery.

The simplest form of DAH7PS is best illustrated by the crystal structures of *Pfu*DAH7PS and *Ape*DAH7PS from the Type Iβ subtype, where an uninterrupted DAH7PS catalytic core is revealed free of any structural decorations.[71-73] The structure comprises a classic $(\alpha/\beta)_8$ TIM barrel fold with the β2-α2 loop capping the entrance of the active site on top of the barrel. The lack of additional structural elements or a discrete regulatory domain means that the enzyme is unresponsive to any downstream metabolites, and hence has unregulated catalytic activity.[71-72]

The rest of the Type Iβ family contains a discrete ACT or CM regulatory domain at either N- or C-terminus of the DAH7PS barrel (Chapter 1.4), and appears largely tetrameric in solution and in crystalline form.[63, 74, 77-78] The addition of these regulatory domains is responsible for feedback regulation of DAH7PS by aromatic amino acids, chorismate or prephenate, respectively. The enzyme *Tma*DAH7PS is a well characterised example of Type Iβ DAH7PS with an N-terminal ACT domain. Presence of the allosteric ligand Tyr or Phe promotes dimerisation of the two ACT domains from opposing chains and effectively blocks the entrance

to the active site (Figure 1.5).[63] A similar gating mechanism of allostery has been reported in *Gsp*DAH7PS, which contains a CM regulatory domain at N-terminus of the barrel. Binding of the ligand, prephenate, to the dimeric CM domain on each side of the catalytic core results in a more compact overall structure with the catalytic component tightly sandwiched by the repositioned regulatory domains, hindering substrate access to the active site.[78]

The Type Iα enzymes contain structural modifications to the core barrel and are typically feedback regulated by aromatic amino acids. The inserted β-sheets and the N-terminal extension contribute to the formation of allosteric site. Characterised DAH7PSs from *E. coli*, *S. cerevisiae* and *N. meningitidis* share similar quaternary structures comprised of dimers of dimer forming the homotetrameric arrangement.[60, 70, 84] Different from the aforementioned regulated Type Iβ family, members of the Type Iα family do not display major conformational changes upon ligand binding, instead subtle changes are involved in communicating the allosteric signal following dynamic and entropically-driven mechanisms.[85]

The first structurally characterised Type II DAH7PS is from *M. tuberculosis*.[86] It adopts a similar homotetrameric quaternary arrangement as Type Iα in the form of dimers of dimer. However, the main reason for its low sequence identity with the Type I counterparts is the extensive structural modifications to the core barrel. The *Mtu*DAH7PS is equipped with a large N-terminal extension (three α-helices and a β-sheet) as well as a double α-helix insertion between α2 and β3, which form the allosteric sites for binding of all three aromatic amino acids. Combination of Phe and Trp synergistically inhibits the enzyme activity, and this inhibitory effect is further enhanced by addition of Tyr.[87] Interestingly, this protein forms a non-covalent complex with *Mtu*CM and contributes to a complex network of regulation in both proteins.[30, 88-89] The allosteric mechanism employed by *Mtu*DAH7PS is rather reminiscent of the Type Iα

family, where no drastic conformational change is observed and allostery is communicated through a series of subtle changes in protein dynamics.

1.4. Regulatory modules

Allosteric regulation of DAH7PS in the Type Iβ family, and some members of the Type II family, involves recruitment of regulatory components in the form of a regulatory domain, or a regulatory partner to form protein complex. It appears that small protein units, such as the ACT domain and CM, are frequently involved in this task. In fact, they have emerged as widely distributed regulatory modules in amino acid metabolism.

1.4.1. ACT domain

Allosteric regulation of enzymes in amino acid and purine metabolism has been found to be conferred recurrently by a small (around 60-70 residues) discrete protein unit, known as the ACT domain.[90] It was first recognised and defined in the 1990s after the crystal structure of *E. coli* 3-phosphoglycerate synthase revealed the βαββαβ fold of the regulatory domain responsible for allosteric inhibition by serine.[91] This regulatory topology was believed to be present in three enzymes in amino acid metabolism including aspartate kinase (AK), chorismate mutase, and TyrA (prephenate dehydrogenase).[92] A significant number of ACT domain containing proteins have been structurally characterised, including ATP phosphoribosyl transferase (ATP-PRT) involved in histidine biosynthesis,[93] some AKs,[94] mammalian amino acids hydroxylases,[95-96] and acetohydroxyacid synthases.[97] The widespread ACT domains have evolved to play regulatory roles in a large group of proteins by binding to

various ligands and often interacting with each other and/or with other functional domains, therefore the sequence identity of the ACT domains is in fact very low, despite their similar structural folds and regulatory purposes.[98]

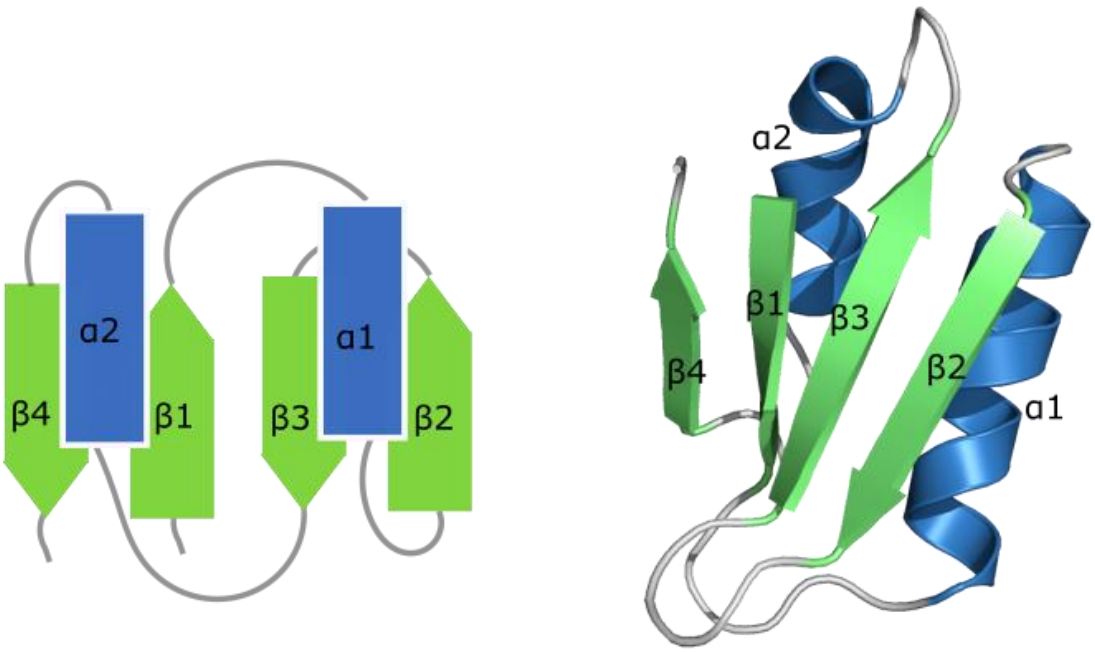


Figure 1.4. The ACT domain from *Tma*DAH7PS showing the typical βαββαβ topology.

The only structurally characterised ACT domain containing DAH7PS is *Tma*DAH7PS. It is a homotetrameric unit with each chain composed of an ACT domain fused to the N-terminus of the catalytic barrel. This ACT domain adopts the typical βαββαβ fold with four antiparallel β-sheets and a pair of α-helices. In the Tyr-bound complex, the four β-sheets are positioned face-to-face to the equivalent β-sheets from the diagonally opposite ACT domain of another chain, forming a dimeric arrangement housing two Tyr binding sites on each side of the tetramer (Figure 1.5). Allosteric ligand binding at the ACT domain interfaces is believed to be associated with changes in protein conformation or quaternary structures in order to transfer signals between the remote regulatory sites and the active sites.[98] Indeed, the differences are clear between the ligand-free and Tyr-bound crystal structures of *Tma*DAH7PS, in which Tyr

binding appears to stabilise the protein in a closed inactive conformation (Figure 1.5). Similar arrangement of the ACT domain has also been identified in the thiamine-binding ACT domains of the thiamine/HMP-binding proteins in *B. subtilis*.[99]

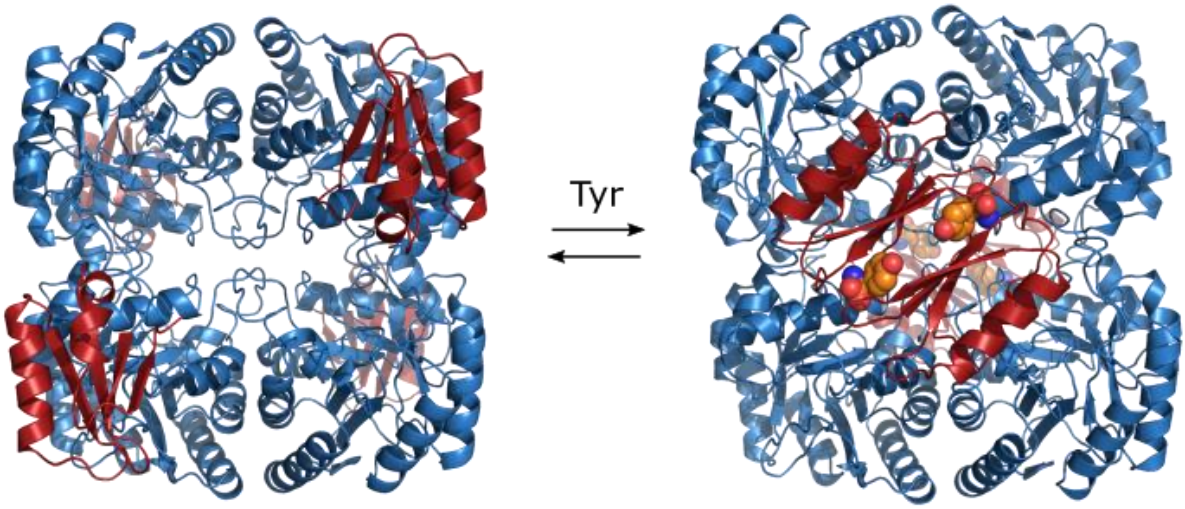


Figure 1.5. Binding of allosteric ligand to the ACT domain (red) promotes large conformational changes in *Tma*DAH7PS (PDB 1RZM left, 3PG9 right). The ligand, Tyr, is shown as orange spheres.

1.4.2. Chorismate mutase

Chorismate as the product of the shikimate pathway is situated at a pivotal metabolic branch point. The conversion of chorismate to prephenate catalysed by CM directly leads to biosynthesis of Phe and Tyr, whereas four other chorismate-utilising enzymes lead the flux towards biosynthesis of other metabolites, including Trp (Figure 1.2). The enzyme CM, as the best-characterised of the chorismate utilising enzymes, has attracted extensive research interest since it was first identified in 1965.[100] This is not only due to its highly efficient catalysis of the unique pericyclic Claisen rearrangement reaction rarely found in enzyme-catalysed reactions (Appendix A), but also because of its wide distribution as a frequently used dual-

functional (catalytic and regulatory) protein module.[27] This type of CMs are mostly homodimeric with each chain composed of three α-helices connected by two turns (Figure 1.6), as illustrated by the example first observed from *E. coli*.[101]

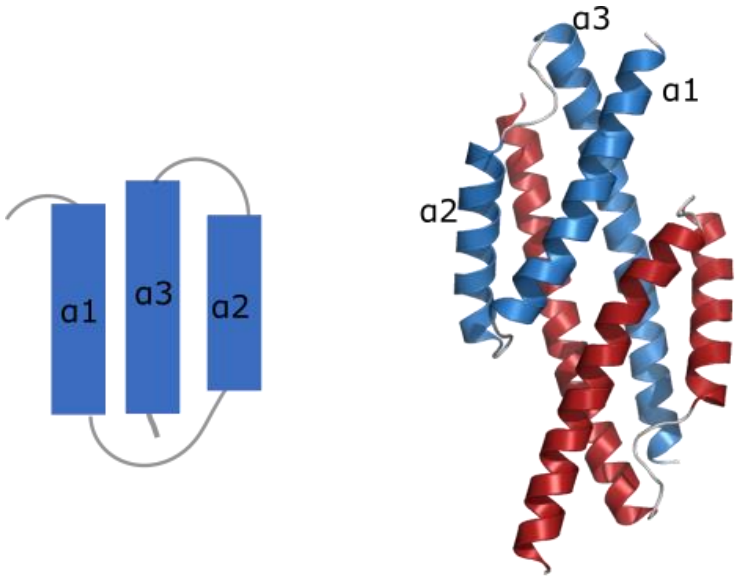


Figure 1.6. The typical three-helical CM architecture represented by *Eco*CM (PDB 1ECM). This type of CM is the most widespread and often adopts a homodimeric quaternary structure (each monomer is coloured separately in blue and red).

Members of the Type Iβ DAH7PS family have developed allosteric regulation in response to chorismate and/or prephenate by recruiting a CM domain.[74-75, 77-78, 102] CM also appears to be involved in allostery of some Type II DAH7PSs (Figure 1.7). *Mtu*DAH7PS forms a protein complex with one CM dimer on each side.[30] Despite the non-covalent nature of the interaction, the structural organisation of the protein complex is somewhat comparable with the Type Iβ DAH7PS that contains a CM domain. Both assemblies comprise the DAH7PS tetrameric core sandwiched by CM dimers. However, in contrast to delivering regulation to the DAH7PS activity in Type Iβ DAH7PS, the recruitment of CM in the *Mtu*DAH7PS-CM complex is involved in a more multifaceted regulatory network, where CM activity is also regulated.[30]

This complex formation allows tight control of flux through two important pathway checkpoints by responding to the level of multiple pathway products.

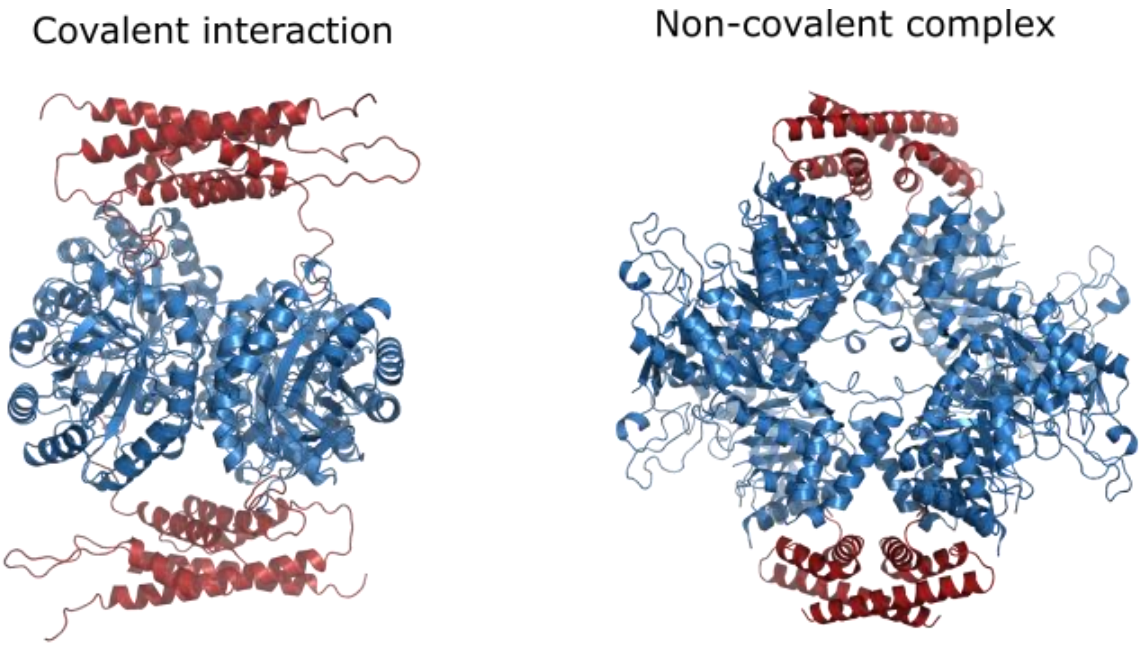


Figure 1.7. CM (red) is involved in allosteric regulation of DAH7PSs via covalent (Type Iβ, *Gsp*DAH7PS) or non-covalent interactions (Type II, *Mtu*DAH7PS).

In addition to its regulatory role in the DAH7PS family, the CM domain has also been identified in some other enzymes involved in the aromatic amino acid metabolic pathways, such as in prephenate dehydratase (PDT) and prephenate dehydrogenase (PDH).[100, 103] The PDT protein, also known as the P-protein, consists of a CM domain fused to a PDT domain, which converts the product of CM (prephenate) to phenylpyruvate. This bifunctional protein has been identified and characterised from a range of Gram-negative bacteria, such as *E. coli*,[100] *Pseudomonas stutzeri*,[104] *Erwinia herbicola*,[105] and *Xanthomonas campestris*.[106] In *Eco*PDT, the CM domain is essential for the cooperative binding of the allosteric ligand Phe to the regulatory domain and prephenate to the PDT domain.[107] On the other hand, the PDH protein

contains a CM domain linked to the PDH domain and is important in the biosynthetic route of Tyr. The PDH domain converts prephenate to 4-hydroxypheylpyruvate, which is subsequently transformed to Tyr. Although this protein is believed to be feedback inhibited by Tyr, the detailed mechanism of regulation remains controversial and variations in the regulatory patterns of the proteins from different organisms are also notable.[108-113]

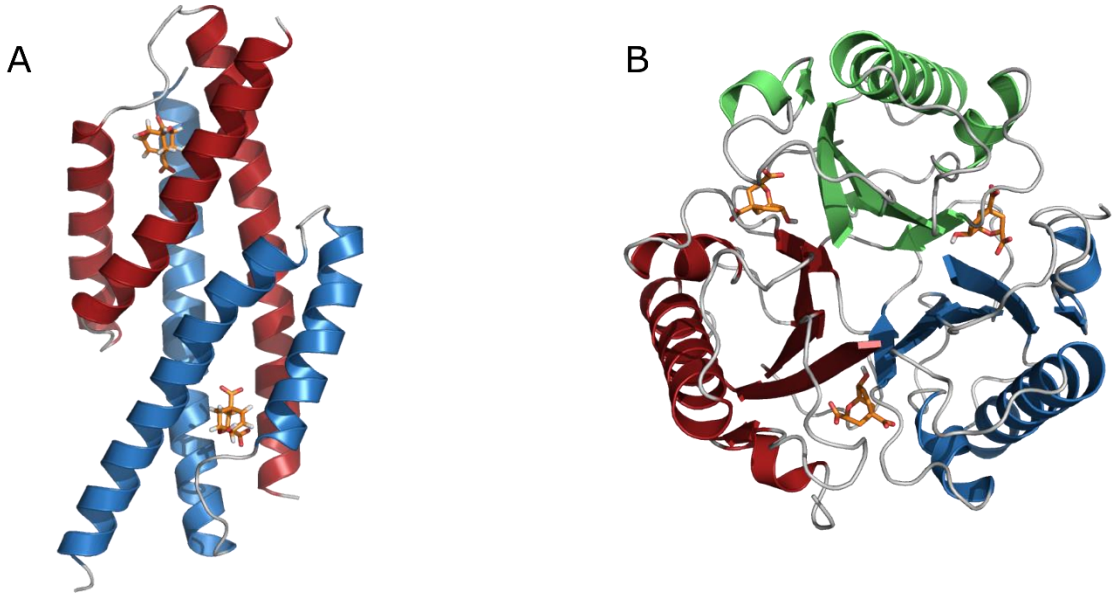


Figure 1.8. Structure of AroQ and AroH CMs. A. *E. coli* CM (PDB 1ECM) of the AroQ class. B. *B. subtilis* CM (PDB 2CHT) of the AroH class. Each chain is coloured separately and the transition state analogue at the active site is shown as orange stick model.

It is worth noting that besides the aforementioned CM, which often acts as regulatory modules, another unrelated structural scaffold of CM also catalyses the same reaction with generally similar efficiencies.[114] In microorganisms, the CM enzymes are divided into two classes, AroQ and AroH (named after the gene), based on differences in structure.[115] In contrast to the previously described homodimeric AroQ CM, the AroH CM is relatively rare in nature and is an independent enzyme characterised by a trimeric pseudo α/β barrel structure with active sites

located at the monomer interfaces (Figure 1.8). This fold was first exemplified in *B. subtilis*.[116] The distinct three-dimensional structures but similar function suggest that the two classes likely arose from convergent evolution.

1.5. Aims of this thesis

The primary aim of this thesis is to provide new insights into the diverse mechanisms of allosteric regulation using a number of examples from the DAH7PS family. As reviewed above, the extraordinary diversity in allostery of DAH7PS has attracted wide research attention for the past few decades, contributing to the understanding of the molecular details governing this process and to its applications in the field of bioengineering and medicine. However, many important questions remain to be answered.

Chapter 2 describes allosteric mechanism that involves large structural changes and answers the questions of whether the conformational change is a general regulatory strategy utilised by the Type Iβ family, and what determines the binding of allosteric ligand in DAH7PSs that contain ACT domain. Functional and structural characteristics of two ACT domain containing DAH7PS enzymes and their variants are described, and molecular determinants for ligand selectivity and potency are determined.

Chapter 3 describes studies on allosteric regulation of DAH7PS enzymes fused with the ACT domain or the CM domain, and provides insights into questions around the relationship between the two types of regulation meditated by the discrete domains. This work demonstrates the interchangeability between different regulatory strategies in the Type Iβ family by

engineering of protein chimera with exchanged regulatory domains. The role of the regulatory domains in conferring allostery and protein oligomerisation is also discussed.

Chapter 4 focuses on exploring the allosteric regulation of Type II DAH7PS from human pathogen *Helicobacter pylori* and another important metabolic enzyme *Hpy*CM. This chapter reveals a number of systematic differences between Type Iβ and Type II enzymes, and answers the question of whether protein complex formation is a common allosteric phenomenon observed in the Type II family. Both biochemical and structural features of the two enzymes are described in relation to the protein complex identified in *M. tuberculosis*.

Given the limitation of current techniques commonly used to understand allostery, **Chapter 5** explores other methods that can access more information on the dynamics of allostery. Specifically, the applications of Förster resonance energy transfer (FRET) and electron paramagnetic resonance (EPR) based techniques are described to answer questions related to the timescales and the delicate molecular motions associated with allostery.

Chapter 2. Allostery and ligand binding in ACT domain-containing DAH7PS

Preface

The ACT domain is a widely distributed ligand binding module often combined with other proteins or domains to confer regulation to its fused partner.[98] DAH7PS with fused terminal ACT domain represents a major class of Type Iβ DAH7PS enzymes. DAH7PS from *T. maritima* (*Tma*DAH7PS) is the only ACT-DAH7PS structurally characterised to date. The crystal structure of *Tma*DAH7PS reveals key interactions at the allosteric inhibitor binding site. Residue Ser31 forms important contacts with the inhibitor, Tyr, which promotes conformational rearrangement to confer allosteric regulation.[63] Although there is overall low sequence similarity in the ACT domains, multiple sequence alignments of ACT-DAH7PSs reveal high level of conservation in the ligand binding region. The motif His-X-Ser, present in *Tma*DAH7PS, is frequently identified in many organisms. Interestingly, another motif His-X-Ile at the equivalent position appears equally abundant. Two protein candidates were chosen that naturally contain the His-X-Ser or the His-X-Ile motif. Characterisation of the two wild-type proteins and their variants was performed to probe the role of the two motifs in allosteric ligand binding.

2.1. Introduction

2.1.1. Ligand binding

In contrast to other DAH7PS enzymes, in which the allosteric ligand binding site is pre-formed by structural extensions appended on the $(\beta/\alpha)_8$ barrel (Chapter 1), allosteric regulation of the ACT domain-containing DAH7PSs is mediated by discrete ACT domains which undergo large repositioning to establish the allosteric site in the presence of ligand.[63, 117-118] Ligand binding in the cavity of the ACT domain interface physically obstructs substrate access to the active site at the C-terminus of the barrel and encumbers catalysis (Figure 1.5).

The only structurally characterised ACT-DAH7PS reported to date is *Tma*DAH7PS. Comparison between the apo (PDB 1RZM) and Tyr-bound (PDB 3PG9) structures revealed the architecture of the cavity formed between two ACT domains from the diagonally opposing chains and uncovered residues that are important for ligand binding (Figure 2.1). The antiparallel β-sheets of ACT domains adopt a face-to-face dimeric arrangement with Tyr bound at the interface. Important hydrogen bonding interactions between the ACT domains and the linker regions include His29 and Ser31, Glu35 and Arg36, as well as Arg36 and Glu75. The ACT domain also makes several contacts with the C-terminus of the adjacent catalytic barrel through hydrogen bonding. In addition to these direct domain interactions in the Tyr-bound structure, the two Tyr molecules contribute key interactions with the ACT domains at each side of the tetrameric protein. The carboxylate functionality of Tyr forms hydrogen bonds with the main-chain amide of Gln34, Glu35 and Arg36 situated at the β2-β3 loop of one ACT domain, and Gly43 of the opposing ACT domain. The amino functionality of Tyr forms hydrogen bonds with the main-chain carbonyl oxygens of Ile41, Gly43 and Asp45. Hydrophobic side-chains of

Val38, Ile42, and Val65 of the opposing domain are also in close contact (<4 Å) with the phenyl ring of Tyr. Residue Ser31 also provides important hydrogen bonding contacts with the phenolic hydroxyl group of the Tyr.

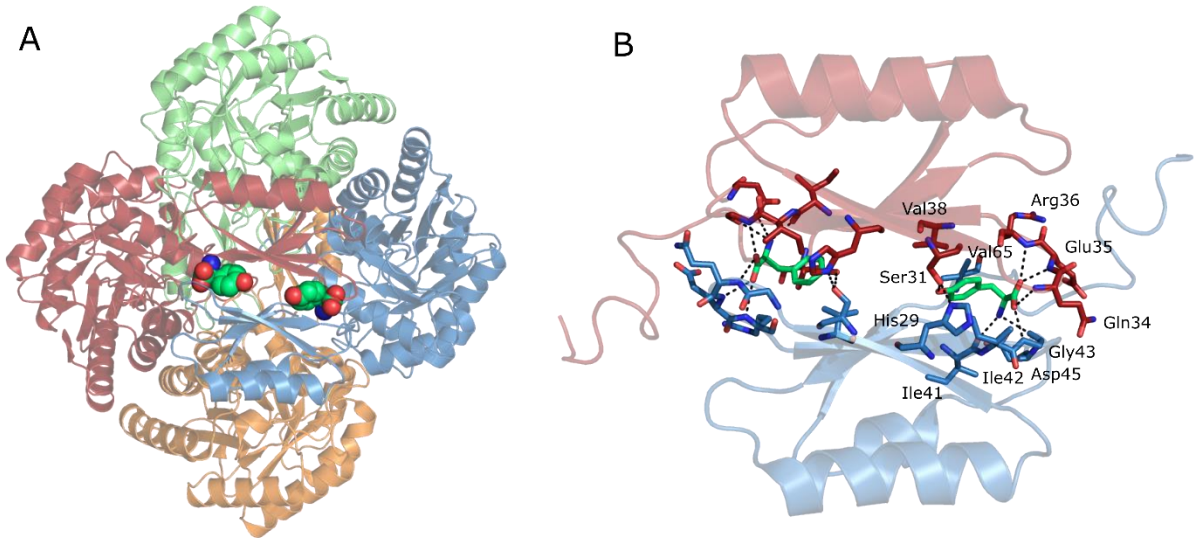
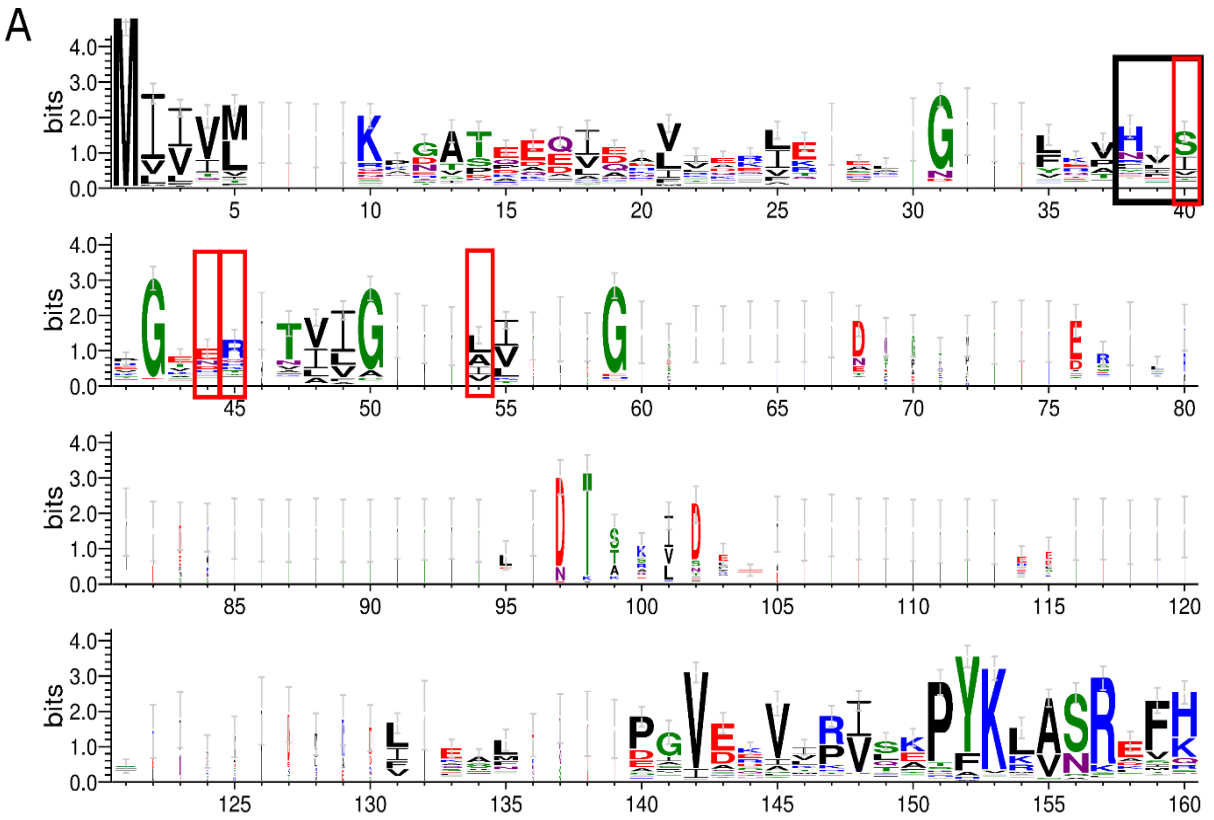


Figure 2.1. Structure of *Tma*DAH7PS with Tyr bound in the cavity between the regulatory domains. A. Top view of the *Tma*DAH7PS tetramer with two Tyr molecules (lime spheres) bound on each side. B. Details of the Tyr (lime sticks) binding sites sandwiched between the regulatory domains of opposing chains. Key residues are labelled and shown as sticks.

2.1.2. Sequence analysis

Previous study of Type Iα DAH7PS from *E. coli* suggests that the level of sequence conservation at allosteric ligand binding region can aid in accurate prediction of key residues involved in ligand sensitivity.[62] Sequence analysis of available DAH7PSs suggests the significant occurrence of DAH7PS with N-terminal ACT-like domain. A search using sequence of the ACT domain and the adjacent linker region (residue 1-80) from *Tma*DAH7PS

identified 197 DAH7PS sequences that share this domain architecture. In general, the important contacts identified from the Tyr-bound crystal structure are not well conserved. For instance, residues involved in direct contact with Tyr such as Glu35, and Asp45 are only very loosely conserved (Figure 2.2 A). However, substantial conservation is observed in the region of His29 to Ser31 (*Tma*DAH7PS numbering), indicating the presence of common motif His-X-Ser, or His-X-Ile at equivalent position (where X is Val, Leu, or Ile, Figure 2.2 B). High level of conservation of this motif is indicative of its important role in ligand binding and specificity. Indeed, a previous study suggests that mutation of some residues around this region impacts severely on ligand potency and specificity.[119] Additionally, the important hydrogen bonding contact formed between the Ser residue and the hydroxyl group of Tyr in the crystal structure suggests that this residue likely plays a critical role in ligand binding, as Phe, which lacks the hydroxyl group, was shown to have decreased inhibitory potency compared to Tyr.[63] In contrast to the direct interaction between Ser31 and Tyr, the conserved His29 is from the opposing chain and forms a hydrogen bond with Ser31, which appears to secure the correct positioning of Ser31 and/or the ligand. Together, it is reasonable to speculate that substitution of Ser to Ile in many organisms, as identified from sequence analysis, may favour binding of Phe as opposed to Tyr due to the increased hydrophobicity. This chapter investigates whether the substitution of a single amino acid residue is sufficient to determine the different allosteric solution of ACT-DAH7PS via Tyr or Phe.



B

```
                                1         10        20        30        40
Thermotoga_maritima             MIVVLKPGSTEEDIRKVVKLAESYNLKCHISKGQERTVIG
Rhodopirellula_maiorica         MIVILQSNVTEEQIDHVLQRIEALGLQHHLSRGTFRTIVG
Rhodopirellula_europaea         MILILKNGVTDDQIDHVIQKVESMGFQTHLSRGTFRTIIG
Gemmatimonas_aurantiaca         MIIVTTATISADALDRIIEHVETAGLRTHVSRGEHRTIVG
Hippea_maritima                 MIIIMKRSATDDDIERVNKKIEKMGLKAHISKGSSKTVIG
Desulfurispirillum_indicum      MVIVLKATATEEDREHVCERVVAFGFRPHVIVGEERTIIG
Geovibrio_sp                    MIVVMRMDAGKDALEEVVEKAEELGFSPHVIFGKERNVIG
Desulfuromonas_acetoxidans      MIIVMEQSSDPKLTDAVVTRIRELGYEPHLIQGSNRQVIG
Thermodesulfovibrio_yellowstonii MVIVMKPEATAEQLQEVISKIEELGYKPHVIYGTTRSVIG
```

Figure 2.2. Sequence conservation in ACT domain of Type Iβ ACT-DAH7PS. A. Region of interest is highlighted with black rectangle. Positions corresponding to the key residues involved in direct hydrogen bonding with Tyr in *Tma*DAH7PS are highlighted in red rectangles. WebLogo3 was used to generate this figure.[120] B. Partial sequence alignment of ACT-DAH7PS showing observed sequence variations at the Tyr-binding position (black rectangle).

2.2. Choice of mutants

To investigate the role of Ser and Ile in ligand specificity of ACT-DAH7PS, two candidates were carefully selected, including one with the His-X-Ser motif and the other with the His-X-Ile motif. As *Tma*DAH7PS is the only structurally characterised ACT-DAH7PS reported to date and has been well studied, it was chosen to examine the role of His-X-Ser. To minimise irrelevant variables, the most similar sequence that comprises the His-X-Ile motif was selected by Blast of *Tma*DAH7PS against all other ACT-DAH7PS sequences. The ACT-DAH7PS from *Thermodesulfovibrio yellowstonii* (*Tye*) was selected, which shares 52% sequence identity with *Tma*DAH7PS (Appendix B, Figure S2).

*Tma*DAH7PS is naturally inhibited by Tyr and to a reduced level by Phe.[63] Mutations were designed on the basis of the sequence conservation, side-chain interactions with Tyr, and the by analogy with a previous study of Type Iα DAH7PS, in which a single amino acid appears to determine ligand binding.[62] Substitution of Ser31 to an Ile in *Tma*DAH7PS was designed to diminish hydrogen bonding capabilities of the ligand binding site with the hydroxyl group of Tyr. Another mutation of Ser31 to a Val was created, as Val was the third most conserved residue at this position after Ser and Ile. Additionally, the degree of hydrophobicity required for potential Phe binding could also be examined with this mutant. Features of the wild-type *Tye*DAH7PS were characterised for the first time and the corresponding Ile to Ser mutation was designed to introduce Tyr binding ability.

2.3. *Tma*DAH7PS mutants

2.3.1. Preparation of *Tma*DAH7PS S31V and S31I

Single-point mutants of *Tma*DAH7PS S31V and S31I were created using QuikChange® site-directed mutagenesis technique with wild-type *Tma*DAH7PS plasmid as template (Chapter 7). The PCR product was treated with DpnI to eliminate contamination from the wild-type plasmid. Once DNA sequence was verified to incorporate the correct mutation, the plasmid was transformed, expressed and purified following the methods for wild-type *Tma*DAH7PS as described fully in Chapter 7.[63] Purified products were confirmed to have the correct molecular weights by mass spectrometry (Appendix B, TableS1).

2.3.2. Kinetic properties

Kinetic parameters for wild-type *Tma*DAH7PS and the two mutants were determined following published procedures by measuring consumption of PEP (Chapter 7). Both *Tma*DAH7PS mutants were catalytically active and the activities were enhanced as the temperature increased, displaying altered kinetic parameters to those of the wild-type protein (Table 2.1, Figure 2.3, Appendix B, Figure S3). They displayed higher K_M values for both substrates compared to the wild-type and lower turnover numbers (k_{cat}). The overall catalytic efficiencies of the two mutants were also reduced.

Table 2.1. Kinetic parameters of *Tma*DAH7PS variants at 60 °C.*

Protein	K_M^{PEP} (µM)	K_M^{E4P}(µM)	k_{cat} (s^{-1})	k_{cat}/K_M^{PEP} ($µM^{-1}s^{-1}$)	k_{cat}/K_M^{E4P} ($µM^{-1}s^{-1}$)
*Tma*DAH7PS^{WT}	8.4 ± 0.7	15 ± 1	14 ± 0.3	1.67	0.93
*Tma*DAH7PS^{S31I}	34 ± 3	53 ± 4	5.4 ± 0.3	0.16	0.1
*Tma*DAH7PS^{S31V}	20 ± 1	51 ± 3	6.1 ± 0.1	0.3	0.12

* Concentration of PEP was fixed at 215 µM when K_M for E4P was determined. Concentration of E4P was fixed at 310 µM when K_M for PEP was determined.

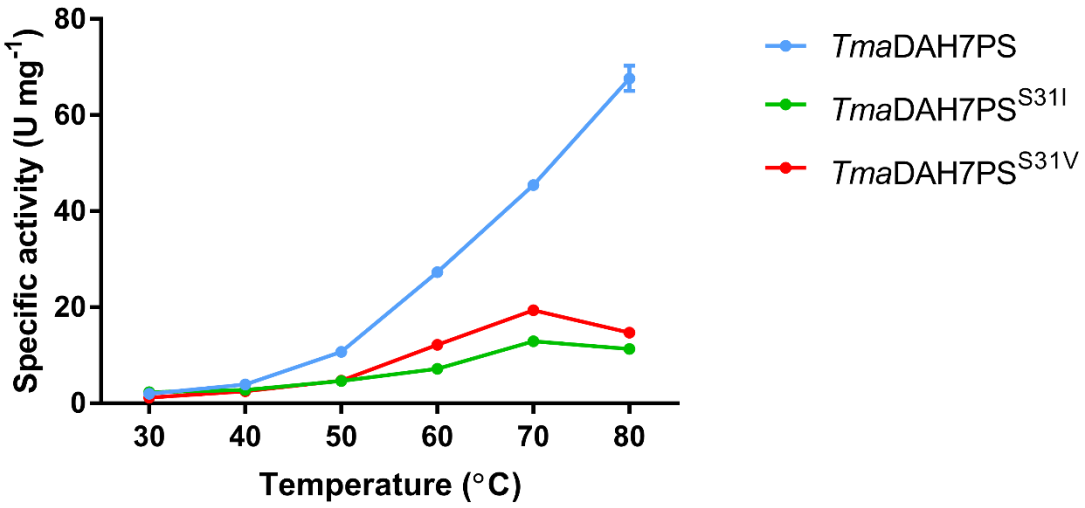


Figure 2.3. Effect of temperature on specific activities of wild-type *Tma*DAH7PS and mutants. Each reaction contained 283 µM PEP, 308 µM E4P, and 100 µM Mn^{2+} in 50 mM BTP with pH fixed at 7.5 at all temperatures tested. Enzyme activities are specified in U (1 U = consumption of 1 µmol substrate·min^{-1}). Specific enzyme activities are given as U·mg^{-1}.

2.3.3. Inhibition

Inhibitory response

The effect of Tyr or Phe on the activity of the mutants was measured by kinetic assays and compared to the wild-type *Tma*DAH7PS (Figure 2.4). Notably, substitution of Ser31 to an Ile or a Val resulted in reduced sensitivity of the protein towards Tyr (Figure 2.4 A). This effect was especially profound for the mutant *Tma*DAH7PS^{S31I}, for which the presence of Tyr did not have a significant impact on its activity under the experimented conditions. Although the activity of the other mutant *Tma*DAH7PS^{S31V} was still affected by the presence of Tyr at high concentrations, the IC_{50} value was increased significantly (greater than 600 µM) compared to that of the wild-type protein (20 ± 2 µM).

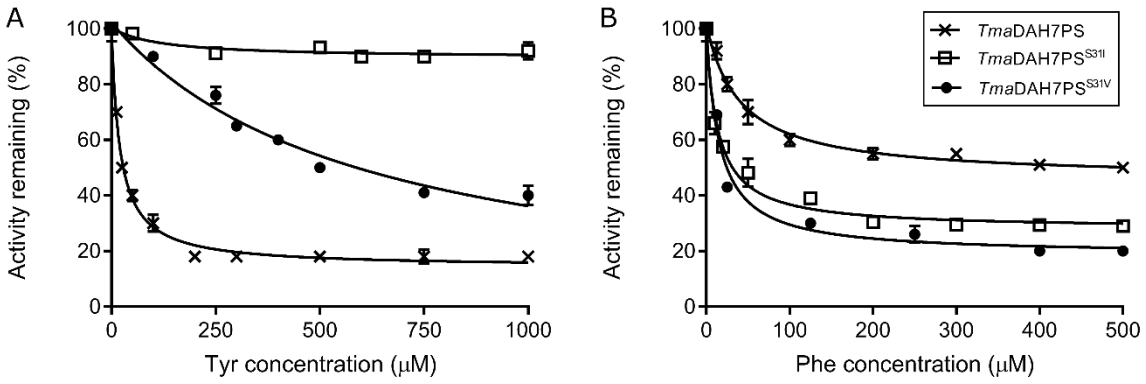


Figure 2.4. Effect of Tyr (A) or Phe (B) on the activity of *Tma*DAH7PS variants at 60 °C. Each reaction contained 283 µM PEP, 308 µM E4P, 100 µM Mn^{2+}, 0 to 1 mM Tyr or 0 to 0.5 mM Phe in 50 mM BTP (pH 7.5). Enzyme was incubated with ligand for at least 5 minutes before reaction was initiated.

In the presence of Phe, the two mutants displayed increased sensitivity and more profound inhibitory response compared to the wild-type *Tma*DAH7PS (Figure 2.4 B). The wild-type protein had approximately 50% activity remaining at high Phe concentrations, whereas the mutants had only 20 to 30% remaining activity. The IC_{50} values for *Tma*DAH7PS^{WT},

*Tma*DAH7PS^{S31I} and *Tma*DAH7PS^{S31V} towards Phe were 41 ± 4, 14 ± 2.4, and 15 ± 3 µM respectively, reflecting the enhanced sensitivity of the mutants towards Phe.

It is worth noting that albeit having a similar trend in ligand sensitivity, the two mutants displayed interesting differences. Mutant *Tma*DAH7PS^{S31I} exhibited clear switched ligand selectivity when compared to the wild-type protein, whereas *Tma*DAH7PS^{S31V} displayed similar degree of inhibition towards Tyr and Phe at high ligand concentrations, although it was still more sensitive towards Phe than Tyr as demonstrated by the IC_{50} values. This difference may relate to the increased hydrophobicity at the ligand-binding site in *Tma*DAH7PS^{S31I} compared to *Tma*DAH7PS^{S31V}. Further study was focused on the *Tma*DAH7PS^{S31I} mutant.

Conformational change

In the presence of the allosteric ligand, Tyr, a large conformational change of wild-type *Tma*DAH7PS, compared to the apo enzyme, was observed by using small angle X-ray scattering (SAXS) experiments (Figure 2.5 A). This technique was also used to examine the solution behavior of the mutant and investigate whether conformational changes could be detected in response to the presence of ligand. SAXS data were collected for the *Tma*DAH7PS^{S31I} mutant in the absence or presence of 1 mM Phe or Tyr.

Initially, *Tma*DAH7PS^{S31I} was injected on the SEC column at high protein concentration (at least 10 mg·mL^{-1}) prior to exposure to X-ray beam. The high concentration was routinely applied due to concerns over signal to noise. However, no obvious change in scattering profile could be observed in the absence and presence of Phe (Figure 2.5 B). This was somewhat surprising as strong inhibition effect of Phe on this mutant was observed by activity assays. Further analysis of the experimental data in comparison with the theoretical scattering calculated from the crystal structures suggested that the protein appeared to be in a ‘closed-

like' conformation under the experimental conditions, regardless of the absence or presence of Phe (Figure 2.5 B).

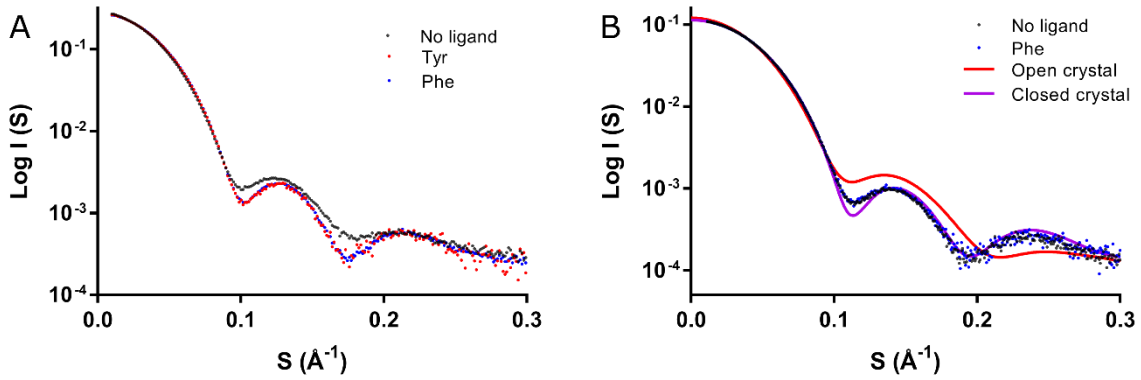


Figure 2.5. SAXS profiles of wild-type *Tma*DAH7PS and *Tma*DAH7PS^{S31I} (10 mg·mL^{-1}). A. The wild-type enzyme displays changes in the presence of Tyr (red dots) or Phe (blue dots), compared to the apo scattering (black dots). B. Scattering profiles and calculated scatterings of *Tma*DAH7PS^{S31I} in the absence or presence Phe. Experimental scatterings are shown as dots and theoretical scatterings are shown as lines. Theoretical scatterings are calculated from crystal structures of wild-type *Tma*DAH7PS (PDB 1RZM and 3PG9). Scattering for *Tma*DAH7PS^{S31I} in the presence of Tyr is not included here, as all scatterings are the same at high protein concentration.

Close inspection of open- and closed-form crystal structures of wild-type *Tma*DAH7PS revealed substantial intermolecular interactions mediated by the ACT domain with the core barrel from adjacent symmetry mates (Figure 2.6). Additional barrel-barrel interactions between different molecules were also present in the crystal structures. At high protein concentrations used for SAXS experiments, intermolecular interactions such as those identified in the crystals were likely enhanced in contrast to a diluted solution. These interactions might affect the average conformation observed from SAXS scatterings, which could contribute to the 'closed-like' conformation observed from the *Tma*DAH7PS^{S31I} mutant.

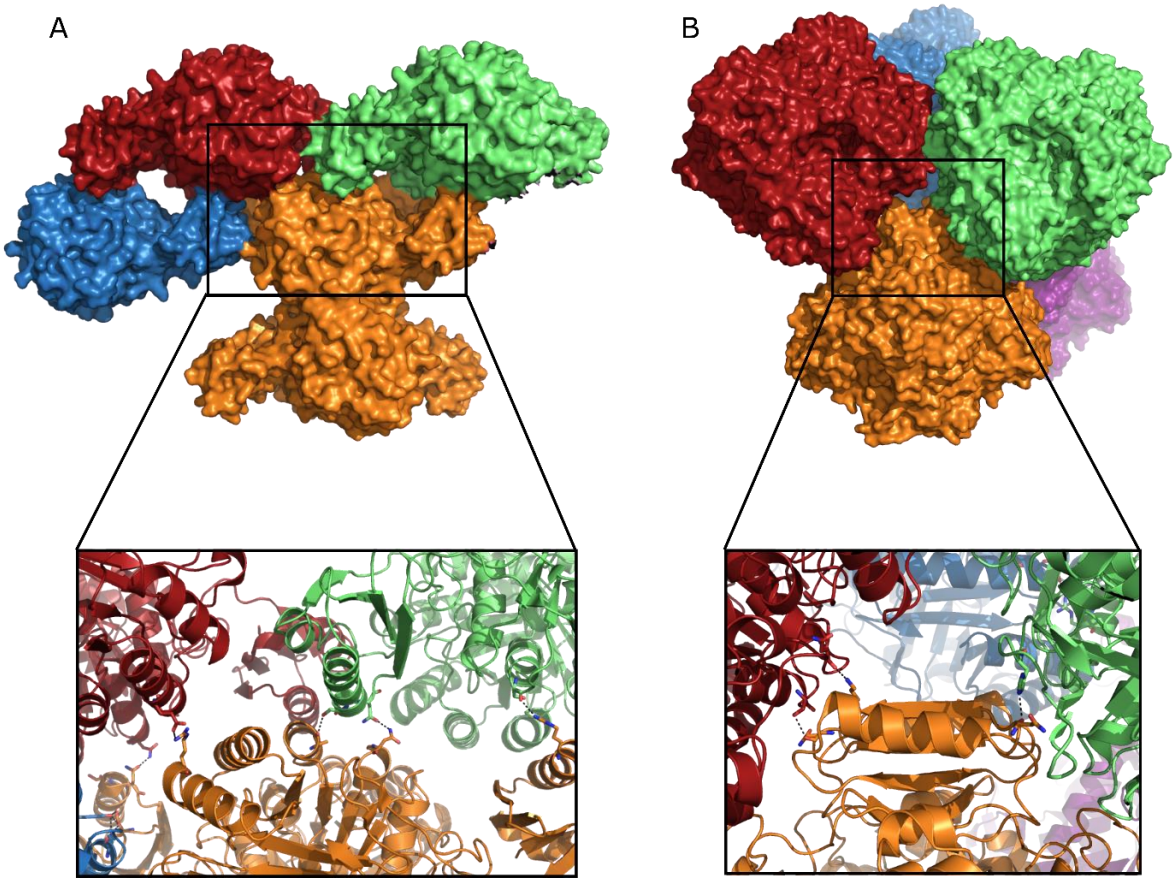


Figure 2.6. Symmetry mates of *Tma*DAH7PS show intermolecular interactions between ACT domain with adjacent barrel, and between barrels. A. Open form. B. Closed form. Bottom panel shows the key residues (sticks) involved in these interactions. Molecule from adjacent tetramer is shown in different colours.

Based on the analyses above, SAXS experiments for *Tma*DAH7PS^{S31I} were repeated at low protein concentrations (1 and 2 mg·mL^{-1}). The SAXS intensity scattering profile can be interpreted by the Fourier-transformation, which transforms I(s) into the P(r)-distribution (Figure 2.7), illustrating the relationship between reciprocal space (Å^{-1}) and real space (Å).[121] The scattering profiles showed clear changes in the presence of Phe at the low concentrations tested (Figure 2.7 A), in contrast to the previous data collected at high concentrations. This data

supports the analysis on intermolecular interactions, and such interactions are reduced for a diluted protein solution. The P(r)-distribution profile is an indication of the data quality and it represents the domain structure and symmetry within the protein molecule so that the initial estimation about the shape of the molecule can be quickly assessed. *Tma*DAH7PS^{S31I} appeared to have an overall globular shape in solution as suggested by the pair distribution plots (Figure 2.7 E), and the presence of Phe shifted the *r* towards smaller distance, indicative of a more compact average state. Indeed, this compaction was reflected in the Kratky plots (Figure 2.7 B). The Kratky plot [$s^2I(s)$ as a function of s] is calculated directly from the scattering curve and provides an excellent way for evaluation of protein folding and flexibility.[122] In the presence of Phe, this mutant appeared less flexible. Although the presence of Tyr seemed to impact on the shape and flexibility of the protein (Figure 2.7 C, D), the degree of observed change was less profound than that caused by Phe.

In addition to the semi-quantitative examination of comparative protein folding, flexibility and conformations from SAXS profiles, the data also allow direct quantitative assessments of the derived structural parameters including radius-of-gyration (R_g), maximum dimension of the particle (D_{max}) and Porod volume (V_p). R_g is defined as the mass distribution of the molecule around its gravity centre, and often used to describe the overall compaction of the molecule. D_{max} measures the longest dimension of the molecule. V_p provides the hydrated volume of the molecule, as determined by application of Porod's law.[121-122] In the presence of Phe, *Tma*DAH7PS^{S31I} appeared to adopt a smaller and more compact state in contrast to the apo state, as suggested by R_g, D_{max} and V_p values (Table 2.2). The presence of Tyr did not cause changes as substantial as that of Phe.

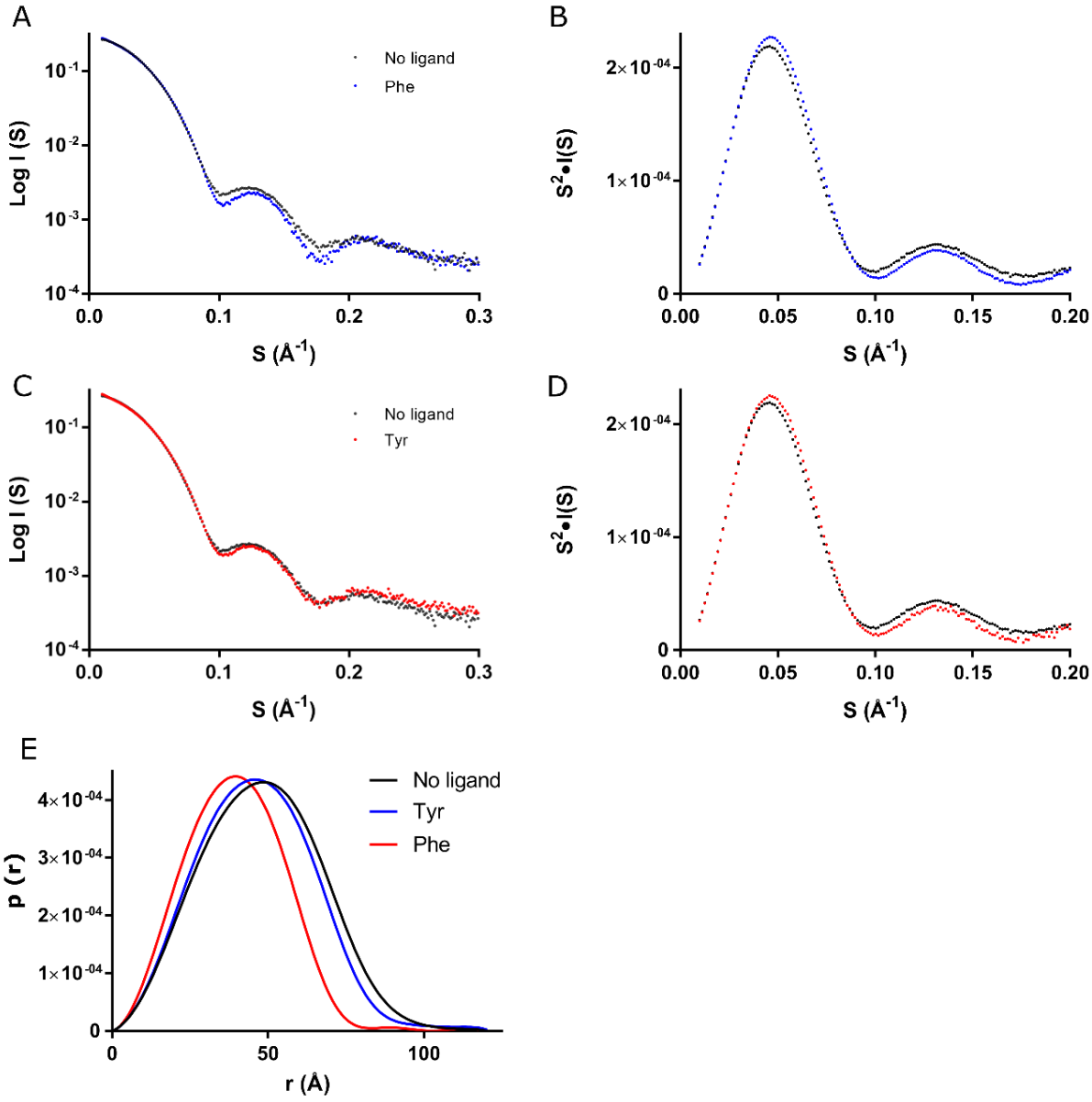


Figure 2.7. A, C. SAXS profiles of *Tma*DAH7PS^{S31I} in the absence (black dots) or presence of Phe (blue dots) or Tyr (red dots). B, D. Kratky plots showing differences in protein flexibility. E. Pair distribution function of the *Tma*DAH7PS^{S31I} in the absence (black line) or presence of Phe (red), or Tyr (blue).

Table 2.2. SAXS parameters of *Tma*DAH7PS^{S31I} in the absence or presence of Phe or Tyr.

Structural parameters	Apo-*Tma*DAH7PS^{S31I}	Phe	Tyr
I(0) (cm^{-1}) [from P(r)]	0.278 ± 0.006	0.261 ± 0.01	0.274 ± 0.006
R_g (Å) [from P(r)]	36.83 ± 0.17	35.89 ± 0.16	36.78 ± 0.15
I(0) (cm^{-1}) (from Guinier)	0.278 ± 0.001	0.261 ± 0.001	0.274 ± 0.001
D_{max} (Å)	146.93	125.09	140.46
R_g (Å) (from Guinier)	36.84 ± 0.2	35.95 ± 0.2	36.78 ± 0.2
Porod volume estimate ($Å^3$)	290,171	289,364	289,854

SAXS data obtained from wild-type *Tma*DAH7PS in the absence of Tyr suggests that the average states of the protein exist in an equilibrium between the open and closed conformations as observed in the crystal structures (Figure 1.5), and the presence of Tyr shifts the equilibrium towards the closed conformation.[63] The mutant *Tma*DAH7PS^{S31I} has increased hydrophobicity at the ligand binding site, which is located right at the interface of the antiparallel β-sheets of two ACT domains. A number of residues with hydrophobic side-chains are found in this region (Figure 2.1 B). It is possible that substitution of Ser to Ile shifts the position of the equilibrium to favour 'closed-like' conformations in the absence of ligand, as shown by the SAXS data.

SAXS experiments were therefore repeated at elevated temperatures with the consideration of changing the position of the equilibrium by increasing the entropic contribution (ΔS) to ΔG (where $\Delta G = \Delta H - T\Delta S$) and to aid the protein in overcoming any local energy barriers. The aim was to investigate whether the equilibrium between different states can be shifted, as well as to examine if the thermophilic nature of the host organism impacts the protein behaviour. Initial tests with a typical SEC-SAXS setup were not ideal due to technical limitations of precise temperature control on the SEC column, and the lag period post SEC and prior to X-ray beam exposure. Therefore, static SAXS experiments were performed for both the wild-type

protein and the mutant on a temperature controlled plate at 25 °C and 50 °C (Figure 2.8). Generally, no significant change was observed as temperature changed. However, it was difficult to conclude the effect of temperature change on the conformations of these proteins from this experiment alone, partially because of the mobile feature of the ACT domain. Additionally, the nature of the static SAXS experiment did not allow separation of populations so that all species in the solution contributed to an averaged scattering profile, which made it difficult to distinguish the population of interest.

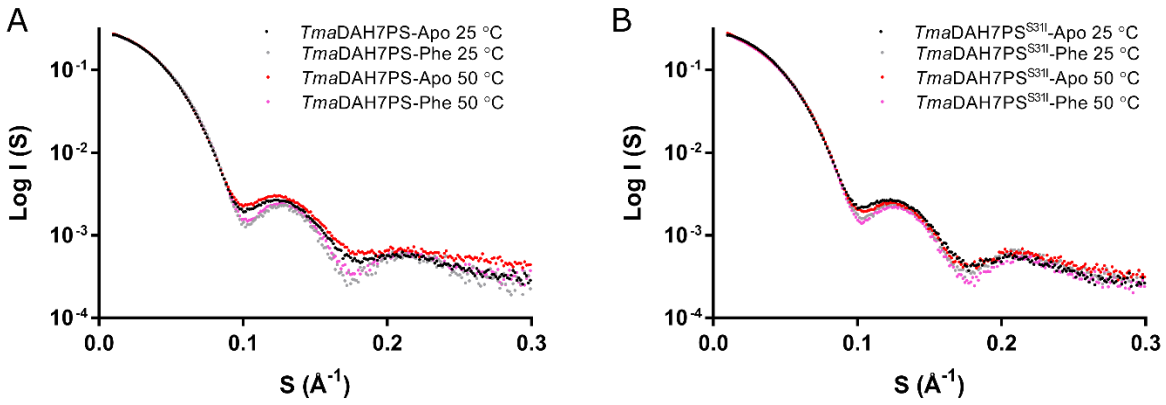


Figure 2.8. Effect of temperature change on SAXS profiles of wild-type *Tma*DAH7PS (A) and *Tma*DAH7PS^{S31I} (B). No significant difference is observed at 25 °C (black/grey dots) and 50 °C (red/pink dots).

Ligand co-crystallisation

Extensive efforts were made in attempt to crystallise the *Tma*DAH7PS^{S31I} mutant in the presence of Phe. Initial sitting-drop crystallisation trials included screens JCSG+, PACT, Clear Strategy I and II screens (Chapter 7). Crystals with various shapes were identified from condition C7 and G5 of the PACT screen, and B8 of the JCSG+ screen. Protein concentration ranging from 9 to 12 mg·mL^{-1} was ideal for crystal growth under these conditions. All three

conditions were optimised and crystals from G5 of PACT and B8 of JCSG+ were examined using the MX2 beamline at the Australian Synchrotron for diffraction and data collection.

Initial diffractions pattern showed heavily smeared spots and were up to 3 Å in resolution (Figure 2.9). Data were processed using XDS,[123-124] cut and scaled using AIMLESS and TRUNCATE (CCP4 program suite),[125] which suggested the absence of strong data at high resolutions with maximum resolution of 4.1 Å, and high mosaicity (1.09). Significant degree of twining was also flagged with 50% twin fraction estimated by L-test and 24% estimated by Britton. Although molecular replacement using MOLREP failed to generate a solution with the wild-type *Tma*DAH7PS crystal structure (PDB 3PG9) as search model,[126] one solution was identified by PHASER, however, it was incorrect as suggested by very poor statistics (R_{free} >45%) from initial refinement. Continuous efforts are being made to obtain better crystallography data.

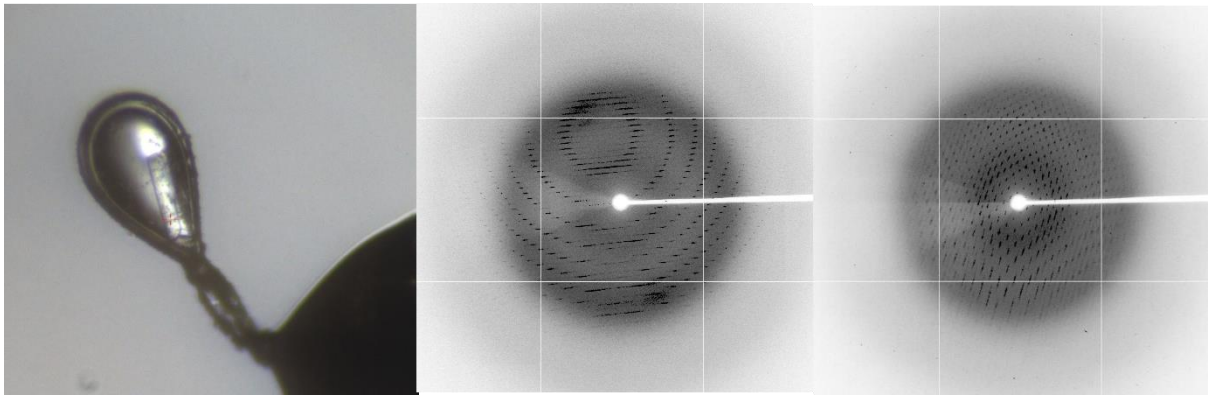


Figure 2.9. Crystal and diffraction images of *Tma*DAH7PS^{S31I} co-crystallised with Phe. Most images show heavy smearing.

2.4. Characterisation of *Tye*DAH7PS

2.4.1. Cloning, expression and purification

The gene encoding for *Tye*DAH7PS was codon-optimised for expression in *E. coli* and purchased from GeneArt. The construct was designed to incorporate a uncleavable C-terminal polyhistidine affinity tag (His_6) preserved from pET28-b(+) vector using NcoI and XhoI restriction site (Chapter 7). This construct was then transformed into *E. coli* BL21*(DE3) cells, for which the expression level was tested to be sufficient. Transformation and expression followed standard procedures described in Chapter 7.

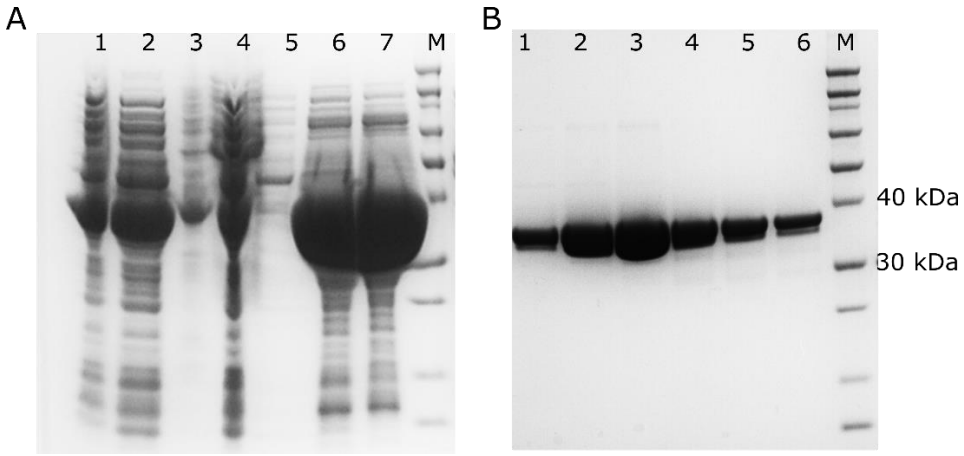


Figure 2.10. Purification of *Tye*DAH7PS. A. 1, crude lysate. 2, soluble fraction. 3, insoluble fraction. 4, unbound fraction from His-Trap column. 5, washed fraction. 6-7, eluted fraction containing protein of interest. B. 1-6 eluted fraction from SEC column. M, protein size marker.

A simple two-step protocol was established for purification of *Tye*DAH7PS (Figure 2.10). The protocol isolated the His_6-tagged protein from crude soluble fraction by using immobilised metal affinity chromatography (IMAC). Desalted protein was then further purified using size-exclusion chromatography (SEC). The yield of the protein was around 45 mg per litre of cell growth. It is worth noting that purification at room temperature was sufficient, as the protein

was heat stable at temperatures up to 60 °C (Figure 2.11), consistent with the observed optimal growth temperature of the organism at 65 °C.[127] An optional heat treatment step at 60 °C could be included before the SEC step, depending on the level of purity.

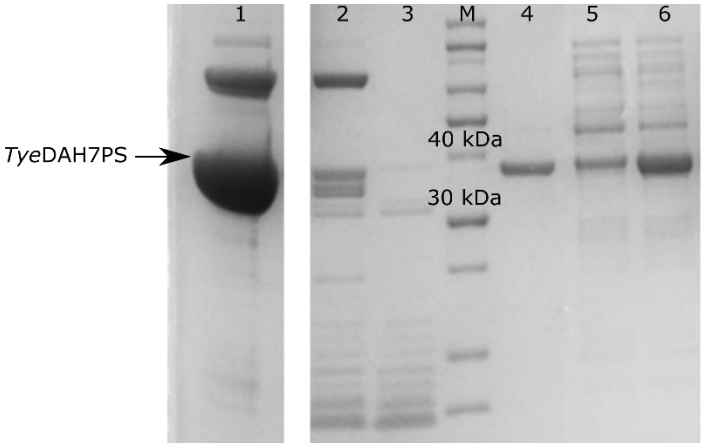


Figure 2.11. Heat stability test of *Tye*DAH7PS. 1-3, soluble fraction at 60 °C, 70 °C, and 80 °C respectively. 4-6, insoluble fraction at 60 °C, 70 °C and 80 °C respectively. M, protein size marker.

2.4.2. Kinetic properties

Metal and temperature dependence

The DAH7PS-catalysed reaction is dependent on the presence of a divalent metal ion as introduced in Chapter 1. The presence of the metal ion in the active site allows activation of E4P as an electrophile, enabling the nucleophilic attack of PEP. A range of divalent metal ions were tested to determine the most activating metal for future experiments (Figure 2.12 A). In the presence of Cd^{2+}, *Tye*DAH7PS was significantly more active than other metal ions tested. The general trend of metal preference largely followed that of *Tma*DAH7PS (Figure 3.3).

The catalytic activity of *Tye*DAH7PS was enhanced as temperature increased (Figure 2.12 B), and diminished at temperatures beyond 70 °C. Compared to the temperature profile of *Tma*DAH7PS (Figure 3.3), this enzyme generally adopted a lower temperature range than that of *Tma*DAH7PS, which was in agreement with the different growth temperatures of the two organisms.

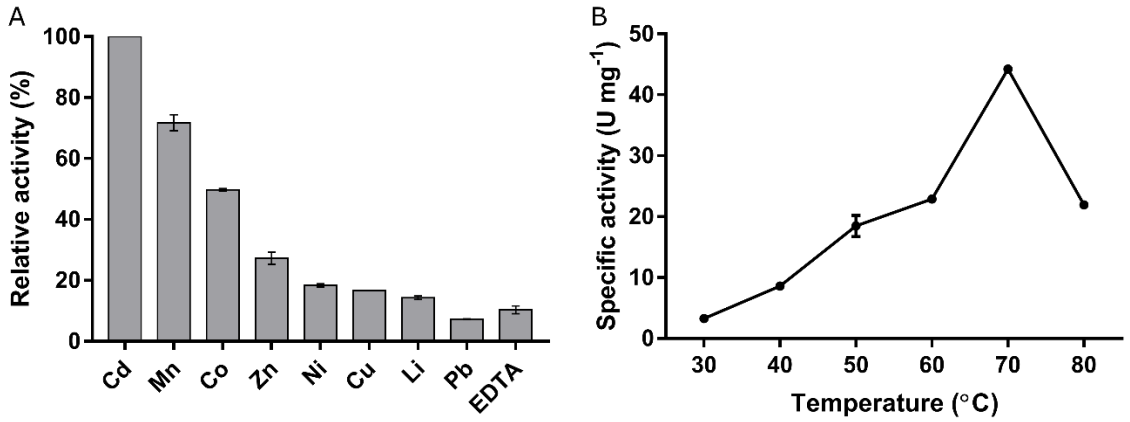


Figure 2.12. Catalytic activity of *Tye*DAH7PS in response to changes in metal ions and temperature. A. DAH7PS activity in the presence of 100 μM metal ion, or EDTA. The mean of triplicate measurements from Cd^{2+} were set as 100% activity. B. The effect of temperature on the DAH7PS activity.

Kinetic parameters

Kinetic parameters for *Tye*DAH7PS catalytic activities were determined following published procedures for *Tma*DAH7PS by measuring consumption of PEP at 232 nm (Chapter 7). Cd^{2+} was used for all assays. The apparent K_M values for PEP and E4P were 55 ± 2.2 μM and 39 ± 2.7 μM respectively, and the k_{cat} value was calculated as 23.8 ± 0.7 s^{-1} (Figure 2.13). These values are generally in line with other Type Iβ DAH7PSs reported in this thesis (Chapter 2.3 and 3.3) with both K_M values in the micro-molar range.

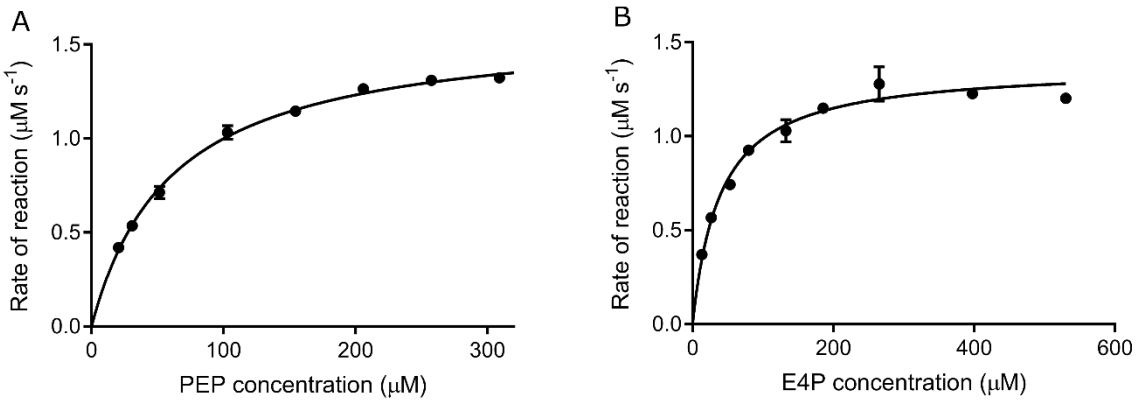


Figure 2.13. Michaelis-Menten kinetics for catalysis of *Tye*DAH7PS with varying PEP (A) or E4P (B) concentrations. Concentration of PEP was fixed at 300 μM when K_M for E4P was determined. Concentration of E4P was fixed at 210 μM when K_M for PEP was determined. Error bar shows SD of triplicate measurements.

2.4.3. Regulation

To test whether *Tye*DAH7PS is naturally sensitive to the presence of Phe and/or Tyr, kinetic assays were performed at 60 °C following the same procedures as that of *Tma*DAH7PS. The catalytic activity of the protein was sensitive to the presence of both Phe and Tyr, and the level of sensitivity was significantly different between the two inhibitors (Figure 2.14). The inhibitory effect introduced by Phe (~12% remaining activity) was substantially more than that of Tyr (~46% remaining activity). The IC_{50} value of *Tye*DAH7PS towards Phe was 25 ± 4 μM, and IC_{50} for Tyr was 121 ± 13 μM. In comparison with the apo-protein (Figure 2.13), substrate affinity was reduced for both substrates in the presence of 500 μM Phe with a K_M value of 176 ± 8.5 μM for PEP and 214 ±11 μM for E4P, resulting in significantly impaired catalytic efficiency.

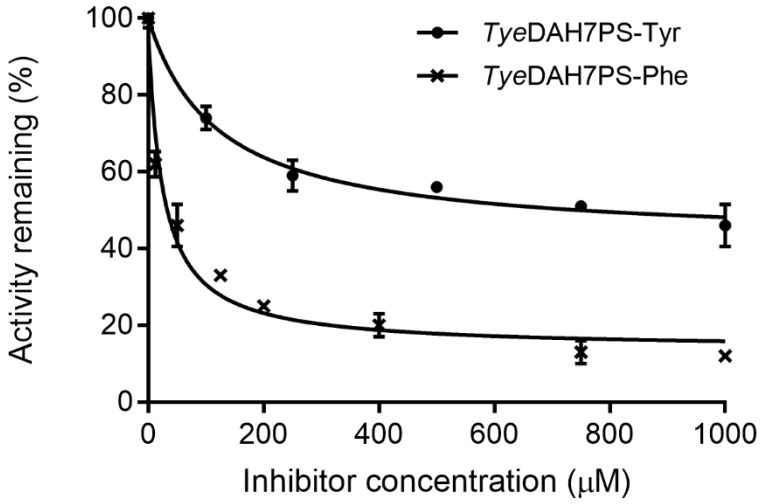


Figure 2.14. Activity of *Tye*DAH7PS in the presence of increasing concentrations of Phe or Tyr. Assays contained 255 μM PEP, 181 μM E4P, 100 μM $CdCl_2$, and 0.05 μM enzyme in 50 mM BTP buffer (pH 7.5 at 60 °C). Error bar represents SD of triplicate measurements.

2.4.4. Structural characteristics

Gel filtration

Structural features of the *Tye*DAH7PS were investigated for the first time. Analytical gel filtration is a standard method used to predict the quaternary structure of a protein in solution based on the molecular size and shape. Compared to advanced biophysical techniques such as AUC, gel filtration is a cheap and simple method widely accessible in many laboratories. Calibration of the analytical SEC column with protein standards of known molecular weight allows the molecular weight of the unknown protein to be determined based on elution volume. The size of *Tye*DAH7PS was determined by extrapolation from a standard curve of the log of molecular weight against elution volume/void volume (Ve/Vo) of protein standards. Different from other Type Iβ DAH7PS enzymes characterised so far, the *Tye*DAH7PS appeared to exist

as tetrameric and dimeric species in solution (Figure 2.15). The calculated molecular weight of each peak of *Tye*DAH7PS was approximately 160 kDa at 1 mg·mL^{-1} injection, in agreement with the size of the tetramer; and 68 kDa, which was smaller than but close to the size of the dimer. Although the C-terminal His_6 uncleavable tag was unlikely to affect oligomeric state of the protein as it was located away from the tetramer interface, another construct without purification tag was created as a control (Table 7.1). This construct also displayed tetrameric and dimeric states and similar kinetic properties compared to the His_6-tagged construct.

The two peaks were collected separately and immediately applied on the analytic SEC column again to assess whether the tetrameric and dimeric species were in equilibrium (Figure 2.15). The protein collected from the tetramer peak (Peak 1) eluted off the column in two peaks, which corresponded to the size of the tetramer and dimer respectively. Protein fractions collected from the dimer peak (Peak 2) was also separated into two peaks with the dimeric species being dominant. This result indicates that *Tye*DAH7PS likely exists in solution in an equilibrium between dimer and tetramer. Interestingly, it seemed that a shift in equilibrium occurred when Phe was added into the two protein samples. Both samples showed a significant increase in tetramer population, indicating the presence of Phe might affect the position of the dimer-tetramer equilibrium and promote the shift of dimer towards tetramer.

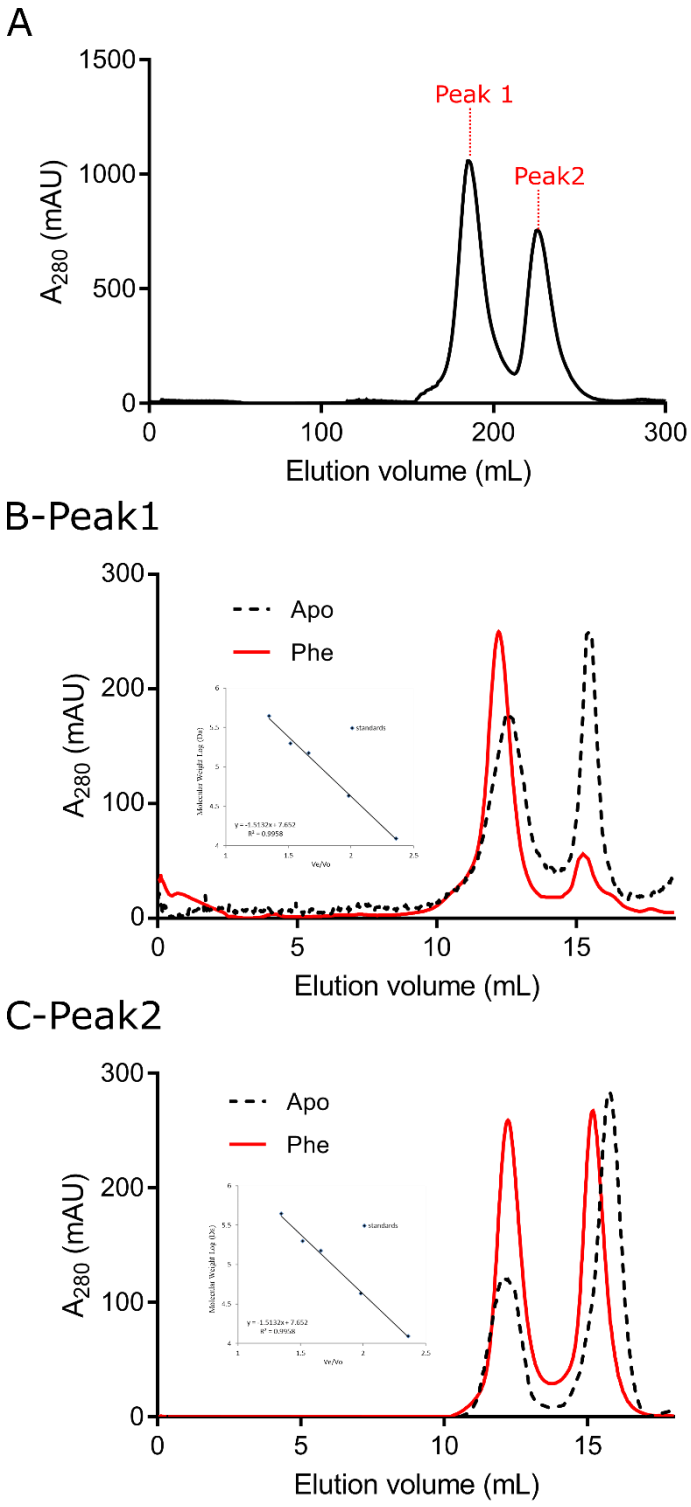


Figure 2.15. Gel filtration of *Tye*DAH7PS in the absence or presence of Phe. A. A typical purification trace of *Tye*DAH7PS contains two major peaks. Peak 1 corresponds to tetramer. Peak 2 corresponds to dimer. B. Elution trace of Peak 1 in the absence or presence of 1 mM Phe at 1 mg·mL^{-1} injection. C. Elution trace of Peak 2 in the absence or presence of 1 mM Phe at 1 mg·mL^{-1} injection.

AUC

AUC experiments were also performed to further examine the interesting observation introduced by the presence of Phe and study the nature of the unique dimer-tetramer equilibrium. In an ultracentrifuge, macromolecular behaviour is observed under the influence of extreme levels of gravity. Under these forces, even small molecules begin to sediment. The sedimentation velocity experiment utilises high velocity and measures the shift of the sedimentation boundary along the radius over time.[128] This movement of molecules through the buffer and the diffusion at the boundary provide information about the state of the molecules as a function of concentration distribution. Operation of the instrument was performed by Dr Sarah Kessans. The AUC sedimentation velocity data was collected for *Tye*DAH7PS in the absence (Figure 2.16 A) or presence (Figure 2.16 B) of Phe at 0.1, 0.4 and 1.6 mg·mL^{-1}, and processed using SEDNTERP and SEDFIT software (Chapter 7).[129]

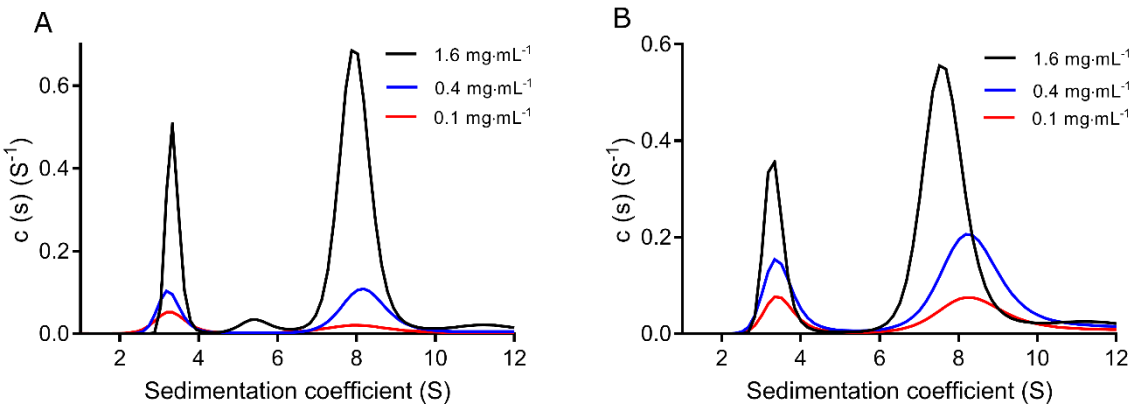


Figure 2.16. AUC sedimentation experiments for *Tye*DAH7PS in the absence (A) or presence (B) of 2 mM Phe. Buffer contained 10 mM BTP (pH 7.5) and 100 mM KCl.

Both data showed the presence of two major oligomers at all concentrations with the small species at approximately 3.4 S and the large species at around 7 to 8 S. The small species corresponded to a molecular weight in between that of a monomer and dimer, the large species

resembled the size of a tetramer. It was difficult to determine whether the smaller species was a monomer or a dimer from this experiment alone as the sedimentation coefficient was affected by not only the molecular weight, but also the shape and density of the molecule.[128] Similarly, the calculated molecular weight from gel filtration experiments in the previous section was also smaller than the theoretical size of a dimer, suggesting the shape of dimer might affect the calculations from gel filtration and AUC velocity experiments. Intriguingly, the presence of Phe appeared to shift the sedimentation coefficient of the tetramer towards smaller values (Figure 2.16 B), and this change was especially notable at higher concentrations. For example, the sedimentation coefficient decreased from 7.9 S (apo) to 7.5 S (Phe) at 1.6 $mg \cdot mL^{-1}$. The tetramer fraction also increased from approximately 65% to 70%. This observation implied that the state of the protein was sensitive to the presence of Phe, which appeared to alter the position of oligomeric equilibrium and make the structure of the protein more compact.

Complementary to the sedimentation velocity experiments, sedimentation equilibrium is a thermodynamic method under smaller centrifugal force, where an equilibrium distribution of species is achieved when the flux by sedimentation and flux by diffusion is balanced.[128] Different from the sedimentation velocity method, sedimentation equilibrium is entirely independent on the shape of the molecule allowing accurate determination of oligomeric state with typically 1-2% precision.[128] The equilibrium constant (K_d) for association of different oligomers may also be derived from the sedimentation equilibrium experiment. This method was used to further investigate the quaternary structure of *Tye*DAH7PS and answer important questions of whether the protein is in a true oligomeric equilibrium, and what the association constant is between different oligomeric states.

The AUC data was collected over a concentration gradient at 7000, 10000, 12000, and 18000 rpm at 20 °C and 280 nm (Figure 2.17). The partial specific volume (v_{bar} = 0.6718 $mL \cdot g^{-1}$),

solvent density ((ρs = 1.0032 g·mL^{-1}) and viscosity (η = 1.00267e-02 cP) were calculated from protein sequence and buffer composition using SEDNTERP.[130] All data obtained from multiple cells and concentrations were modelled globally using SEDPHAT with the standard monomer-dimer association model applied. The *Tye*DAH7PS dimer was defined as the 'monomer' and *Tye*DAH7PS tetramer was defined as the 'dimer'. The monomer-dimer association model produced the best global fitting at all conditions, compared to other models tested including the monomer-n-mer association model and the monomer-dimer-tetramer model. This supports the dimer-tetramer equilibrium of *Tye*DAH7PS and verifies that the small species observed previously from analytical gel filtration and sedimentation velocity experiments is dimeric, rather than monomeric. Key parameters were derived from data collected at 0.2 mg·mL^{-1} for maximum accuracy (χ^2 = 1.357, rmsd error of fit = 0.0028). The oligomeric equilibrium was also reflected by the fitted molecular weights of approximately 79 kDa and 158 kDa, in precise agreement with the theoretical values for the dimer and tetramer respectively. The dissociation constant, K_d, for the dimer-tetramer equilibrium was calculated to be 5.13 µM. The dimer (D)-tetramer (T) association is described as below:

$$D + D \overset{Kd}{\Leftrightarrow} T$$

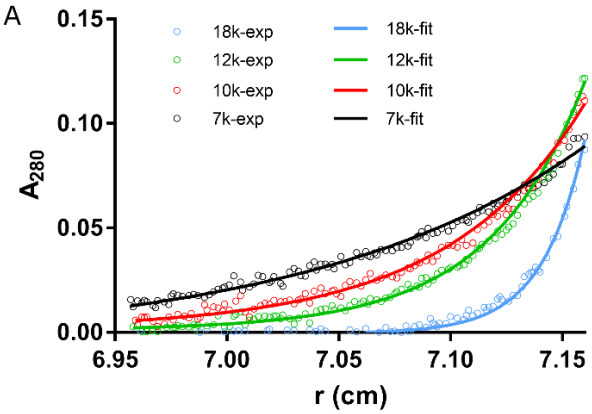
Dissociation constant K_d is defined as:

$$Kd = \frac{[D]^2}{[T]}$$

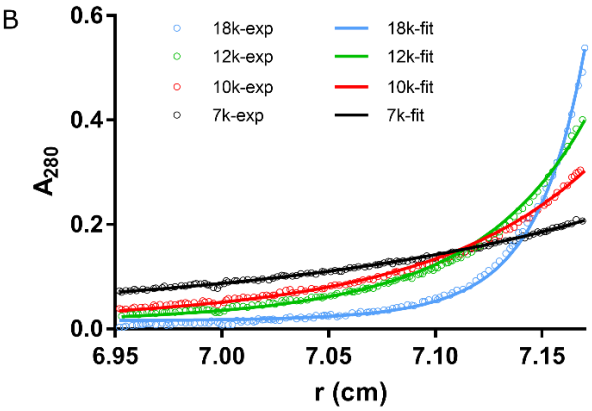
Where [D] and [T] present the molar concentrations of dimer and tetramer respectively. The total concentration of protein dimers can be expressed as:

$$E_{total} = [D] + 2[T]$$

A



B



C

χ^2	1.357
Fitted M_W1 (Da)	79092
Fitted M_W2 (Da)	158184
log10 (Ka_app)	5.290
Fitted Kd (μM)	5.13

Figure 2.17. Analysis of AUC sedimentation equilibrium data for *Tye*DAH7PS. Experimental radical distribution profiles (shown as circles) of the protein at all concentrations [0.1 (A), and 0.2 (B) mg·mL^{-1}] are best fitted to the monomer-dimer model (shown as solid lines) with the 'monomer' defined as *Tye*DAH7PS dimer and 'dimer' defined as *Tye*DAH7PS tetramer. Experiments were performed at 7000 (black), 10000 (red), 12000 (green), and 18000 (blue) rpm, in buffer containing 10 mM BTP (pH 7.5) and 100 mM KCl. Fitting from the 0.2 mg·mL^{-1} data (B) produces the best global reduced χ^2 value and is used for derivation of other key parameters (C).

The dissociation constant (K_d) of 5.13 µM suggests that the protein is likely to be primarily dimeric under kinetic assay conditions in the absence of Phe. Therefore, the strong inhibition in catalysis observed in the presence of Phe may support the role of Phe in altering the position of the dimer-tetramer equilibrium as the tetramer form is expected to be responsible for inhibition. Similar AUC sedimentation equilibrium experiments should be performed in the future to assess in detail the shift in sedimentation coefficient observed from velocity experiments introduced by the presence of Phe, which will aid in defining more clear the effect of Phe on the oligomeric equilibrium.

SAXS

Initially SEC-SAXS was performed in attempt to separate and observe the dimer and tetramer peaks of *Tye*DAH7PS by SEC prior to X-ray exposure. As relatively low protein concentration was required to detect both UV_{280} peaks, it was difficult to achieve sufficient scattering intensity to produce satisfactory level of signal to noise. Static SAXS experiments were then designed with a concentration gradient applied based on the K_d value (5.13 µM) of the apo-protein obtained from AUC, as well as limits on signal to noise. A range of theoretical concentration ratios between dimer (D) and tetramer (T) were calculated and used based on the relevant equations described previously. These [D]:[T] ratios included 2:1, 1.5:1, 1:1, and 1:1.5, with total protein monomer concentrations of approximately 0.5, 1, 2, and 4 $mg \cdot mL^{-1}$.

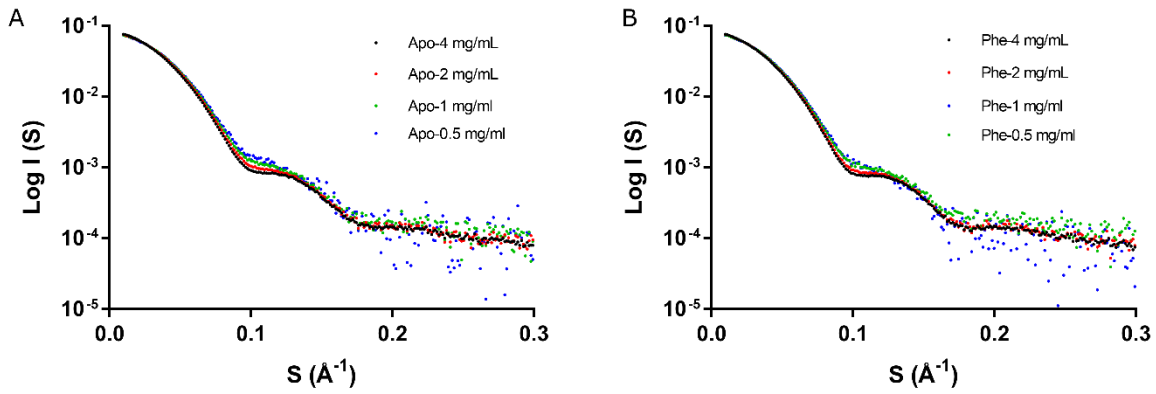


Figure 2.18. SAXS scattering of *Tye*DAH7PS in the absence (A) or presence (B) of 1 mM Phe. Concentration series of 0.5, 1, 2, and 4 mg·mL^{-1} were applied.

Concentration dependent changes in scattering profiles were evident both in the absence and the presence of Phe (Figure 2.18). While in the absence of Phe, a clear trend was observed that the averaged molecule became larger as total protein concentration increased (Table 2.3), the trend in the presence of Phe was not straight forward. As the K_d value used to calculate the theoretical concentrations of dimer and tetramer was derived from AUC of the protein in the absence of Phe, and the effect of Phe on K_d was not yet known, the calculated ratio of dimer and tetramer may not represent the real experimental ratios in the presence of Phe. This unknown equilibrium constant in the presence of Phe posed complications in analysing the SAXS data. As described in Chapter 2.3, the addition of Tyr in *Tma*DAH7PS caused the protein structure to change from an open conformation to a closed and more compact conformation, leading to decrease in SAXS structural parameters such as R_g, V_p and D_{max} for an overall smaller structure. For the homologous *Tye*DAH7PS, similar inhibitory effect occurred in the presence of Phe. Assuming Phe impacted on the *Tye*DAH7PS structure in a similar manner as Tyr did on *Tma*DAH7PS, smaller values of the SAXS parameters were expected in the presence Phe for the tetrameric *Tye*DAH7PS molecules. From Table 2.3, this was the case when the tetramer concentration was high. For instance, values of R_g, V_p and D_{max} were all clearly reduced when

Phe was added to the 1D:1.5T sample. However, as the dimeric species became dominant (i.e. as total protein concentration decreased), there was no clear trend on the changes of SAXS structural parameters. As observed preciously from gel filtration and AUC experiments, the presence of Phe appeared to alter the position of the dimer-tetramer equilibrium and promote the formation of tetramer from dimers. Therefore, in the samples that contained supposedly more dimers, the addition of Phe might cause more tetramers being formed resulting in an increased average molecule size. These tetramers were then subject to inhibition by Phe and became more compact resulting in a smaller average molecule size. This mixed contradictory effect on the overall molecule size may explain the variations in the structural parameters in the presence of Phe at low protein concentrations.

Table 2.3. SAXS structural parameters of *Tye*DAH7PS in the absence or presence of 1 mM Phe. Theoretical ratios represent molar ratios of dimer (D) to tetramer (T) based on the dissociation constant K_d of 5.13 μM. Estimation of sample composition was achieved by using OLIGOMER with *Tma*DAH7PS coordinates as templates.

Theoretical ratio_Apo	**R_g (Å)**	**V_p (Å^3)**	**D_{max} (Å)**	**Dimer (%)**	**Tetramer-open (%)**	**Tetramer-closed (%)**
2D:1T	37.31 ± 0.6	245,477	149.75	47.9 ± 21.7	39.3 ± 20	12.8 ± 1.2
1.5D:1T	38.85 ± 0.5	292,095	150.52	32.6 ± 22.2	57.4 ± 18.6	9.4 ± 4.8
1D:1T	40.26 ± 0.08	342,564	164.52	21.6 ± 10.9	64.4 ± 10.8	13.9 ± 0.8
1D:1.5T	42.12 ± 0.12	392,092	174.04	10.5 ± 8	73.1 ± 5.5	16.3 ± 0.6
Theoretical ratio_Phe	**R_g (Å)**	**V_p (Å^3)**	**D_{max} (Å)**	**Dimer (%)**	**Tetramer-open (%)**	**Tetramer-closed (%)**
2D:1T	38.45 ± 0.7	271,421	136.63	21.5 ± 22.2	27.8 ± 9.2	50.6 ± 20.3
1.5D:1T	38.30 ± 0.7	258,440	140.38	19.6 ± 9.1	29.6 ± 6.6	50.7 ± 10.8
1D:1T	40.29 ± 0.6	303,745	157.16	16.8 ± 5	20.9 ± 2.1	62.2 ± 5.5
1D:1.5T	40.37 ± 0.2	328,634	157.02	9.9 ± 1.9	25.5 ± 0.5	64.5 ± 4.8

OLIGOMER was also used to analyse the composition of each sample (Table 2.3). Coordinates from crystal structures of *Tma*DAH7PS 1RZM, 3PG9 and a dimeric form derived from 1RZM were applied as templates to fit experimental data of *Tye*DAH7PS. A general trend was derived, even though the error appeared large, which was possibly due to the mixed effect caused by the heterogenous samples having both oligomeric and allosteric changes. In the absence of Phe, most tetrameric species were in an 'open-like' form. In contrast, the presence of Phe seemed to reduce the dimer component and the majority of the tetramer species appeared closed, supporting the shift of K_d caused by the presence of Phe. The variations between theoretical ratios calculated based on K_d and experimental ratio estimated using OLIGOMER were also noted. OLIGOMER predicted higher ratio of the tetrameric form compared to the theoretical value. This was partially because the default setting of the program counted aggregations as the largest input molecule. It should be noted that OLIGOMER prediction applied here only represents a general estimation and the real K_d value may also be affected by experimental conditions such as temperature and pH. Future SEC-SAXS experiments at high protein concentrations may be designed to observe the conformational change of tetrameric *Tye*DAH7PS introduced by Phe without interference from the dimeric species. However, the effect of Phe on the dimeric species is difficult to be captured by SEC-SAXS experiments due to limitations of signal to noise as very low protein concentrations are required.

Crystallisation

Continuous efforts were made to crystallise the *Tye*DAH7PS. Several 96-well conditions from Molecular Dimensions were screened using sitting-drop vapour diffusion method, including JCSG+, PACT, Clear Strategy I and II, and Proplex. The protein precipitated easily and its concentration must be limited to lower than 6 $mg \cdot mL^{-1}$ to achieve effective screening. Additives including PEP, E4P, Cd^{2+}, and Phe were also included in the screening individually

or in combinations. Unfortunately, these approaches were unsuccessful in producing lead protein crystals. Purified *Tye*DAH7PS samples in the absence or presence of Phe were then sent to the Collaborative Crystallisation Centre (C3) facility in Melbourne for further screening of crystallisation conditions. Almost 400 conditions were screened and monitored for two and half months, however no lead condition has been identified. The oligomeric equilibrium of *Tye*DAH7PS may contribute to the difficulty in crystallisation.

2.5. *Tye*DAH7PS mutant

2.5.1. Preparation of *Tye*DAH7PS I31S

Site-directed mutagenesis was performed to create the single-point mutant, *Tye*DAH7PS^{I31S}, by using QuikChange® technique with wild-type *Tye*DAH7PS-pET28-b(+) plasmid as template (Chapter 7). The sequence of the PCR product was verified to incorporate the correct mutation. The plasmid was transformed in *E. coli* BL21* (DE3) cells. The protein was expressed and purified following the same method developed for the wild-type *Tye*DAH7PS. Purified mutant was confirmed to have the correct molecular weight by mass spectrometry (Appendix B, Table S1).

2.5.2. Kinetic properties

Kinetic parameters for *Tye*DAH7PS^{I31S} were determined following the same procedures as for the wild-type protein by measuring consumption of PEP. The mutant was catalytically active and displayed mildly altered kinetic parameters compared to those of the wild-type protein (Table 2.4, Appendix B, Figure S3), with similar K_M for PEP and slightly increased K_M for E4P. The turnover number was increased, which largely contributed to the slightly enhanced catalytic efficiencies for both substrates.

Table 2.4. Kinetic parameters of *Tye*DAH7PS^{WT} and *Tye*DAH7PS^{I31S}.*

Protein	K_M^{PEP} (μM)	K_M^{E4P}(μM)	k_{cat} (s^{-1})	k_{cat}/K_M^{PEP} ($\mu M^{-1}s^{-1}$)	k_{cat}/K_M^{E4P} ($\mu M^{-1}s^{-1}$)
*Tye*DAH7PS^{WT}	55 ± 2	39 ± 3	24 ± 1	0.44	0.62
*Tye*DAH7PS^{I31S}	56 ± 5	46 ± 4	32 ± 1	0.57	0.69

*Concentration of PEP was fixed at 300 μM when K_M for E4P was determined. Concentration of E4P was fixed at 210 μM when K_M for PEP was determined.

2.5.3. Inhibition

Mutation of Ile to Ser in the regulatory domain was hypothesised to switch ligand selectivity of *Tye*DAH7PS from Phe to Tyr, complementing the ligand switch by S31I mutation in *Tma*DAH7PS. This hypothesis was tested by adding Tyr or Phe in the enzyme kinetic assays for *Tye*DAH7PS^{I31S} (Figure 2.19). The presence of Tyr inhibited the enzyme activity by approximately 80%, while the presence of Phe inhibited catalysis by around 30%, opposite to the inhibitory effect observed for the wild-type *Tye*DAH7PS (Figure 2.14). The IC_{50} values for Tyr and Phe were 73 ± 12 and 161 ± 29 μM respectively. The IC_{50} value of Tyr was reduced by almost two-fold compared to the wild-type protein, suggesting higher sensitivity of the mutant towards Tyr. The IC_{50} value of Phe was increased by almost six-fold, indicating that Phe binding to *Tye*DAH7PS^{I31S} was much more unfavorable.

Similar to the wild-type *Tye*DAH7PS, mutant *Tye*DAH7PS^{I31S} was also in a dimer-tetramer equilibrium (Figure 2.20 A). As a result, similar complications are involved in SAXS analysis. Nevertheless, scattering profile of *Tye*DAH7PS^{I31S} was indicative of some sort of conformational change when Tyr was present (Figure 2.20 B). The overall shape of the averaged molecule appeared to display more globular features in the presence of Tyr compared

to the apo protein. As discussed for the wild-type protein, this conformational change in SAXS profile observed for *Tye*DAH7PS^{I31S} was likely contributed by the combination of a few factors. First, structural changes of the tetramer population from 'open-like' to 'closed-like' conformation may occur upon addition of Tyr. Second, the formation of tetramers from dimers may be facilitated with the presence of Tyr. Additionally, substitution of Ile to Ser at the ACT domain interface may also change the population distribution. Further AUC sedimentation velocity and equilibrium experiments are necessary for determining the dissociation constant of this mutant in the absence or presence of Tyr, so that accurate SAXS experiments may be designed that provide more detailed solution behaviour of the dimeric and tetrameric species.

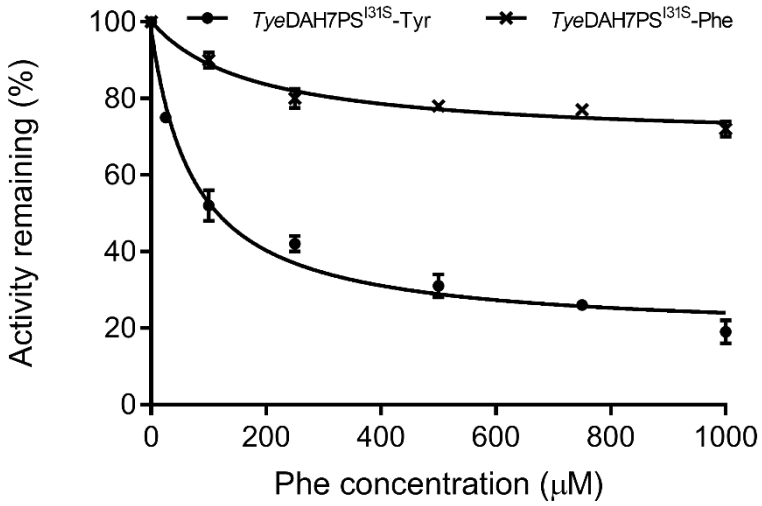


Figure 2.19. Activity of *Tye*DAH7PS^{I31S} in the presence of Tyr or Phe. Assays contained 255 μM PEP, 220 μM E4P, 100 μM $CdCl_2$, various concentrations of Tyr or Phe (0-1 mM) and 0.05 μM enzyme in 50 mM BTP buffer (pH 7.5 at 60 °C). Error bar represents SD of triplicate measurements.

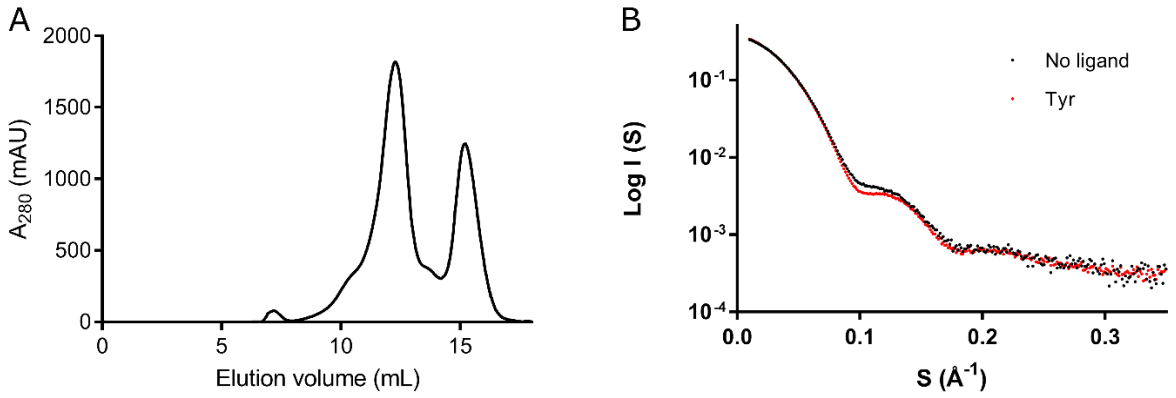


Figure 2.20. A. Gel filtration of *Tye*DAH7PS^{I31S}. Similar to the wild-type *Tye*DAH7PS, this mutant is in a dimer-tetramer equilibrium. B. SAXS scattering of *Tye*DAH7PS^{I31S} in the absence or presence of 1 mM Tyr. Protein concentration applied was around 2 mg·mL^{-1}.

2.6. Summary and discussion

The *Tma*DAH7PS-like Type Iβ DAH7PS enzymes represent a major class of Type Iβ DAH7PS, in which the undecorated $(\beta/\alpha)_8$ catalytic barrel is attached to an ACT-like domain responsible for allosteric ligand binding. Residue Ser31 in the ACT domain of *Tma*DAH7PS is important for Tyr binding, which promotes dimerisation of two ACT domains from diagonally opposing DAH7PS chains and hinders substrate entrance to the active site. In this Chapter, the detailed characterisation of another Type Iβ DAH7PS from *T. yellowstonii* shows that the ACT domain-mediated regulation is likely a general allosteric mechanism utilised by this class of DAH7PS enzymes. The studies performed on Ser and Ile mutations of *Tma*DAH7PS or *Tye*DAH7PS highlight the importance of these residues in determining ligand selectivity and potency.

2.6.1. Inhibition of *Tye*DAH7PS

The *Tye*DAH7PS is functionally similar to *Tma*DAH7PS in many aspects, such as catalytic efficiencies, metal dependencies, and temperature-dependent activity profiles. Instead of being inhibited by Tyr, *Tye*DAH7PS displays strong inhibition in the presence of Phe as predicted based on the substitution of Ser to Ile in the ligand binding site of the ACT domain. Despite the lack of a crystal structure, the inhibition profile and results from AUC and SAXS experiments indicate that the mechanism of inhibition employed by *Tye*DAH7PS likely resembles that of *Tma*DAH7PS, in which a physical gating mechanism occurs in the presence of inhibitor.

Homology models of the open- and closed- forms of *Tye*DAH7PS have been kindly provided by Dr Wanting Jiao using the crystal structures of *Tma*DAH7PS as templates (PDB IRZM and 3PG9). The model allows the prediction of the binding conformation for Phe. Phe binds to the hydrophobic pocket formed at the dimer interface constructed by the antiparallel β-sheets of two ACT domains. In contrast, Tyr binding appears to be relatively unfavourable due to the high hydrophobicity around the location of its hydroxyl functionality. The benzene side-chain of Phe are primarily stabilised by multiple nearby hydrophobic residues located on β2, β3 and beginning of β1 of the ACT domain, including Val2, Ile31, Val38, and Val42 (Figure 2.21). In this model, the carboxylate functionality of Phe can form hydrogen bonds with nearby charged residues such as Arg46 on the α2-β3 loop, and polar residues on the β2-β3 loop from the other chain such as Thr34 and 35. The amino group of Phe can be stabilised by some charged residues from the α2-β3 loop, such as Glu45. In this predicted ligand binding pocket, His29 is the only charged residue positioned near the hydrophobic side-chain of Phe. This residue is likely important in stabilisation of the phenolic hydroxyl group of Tyr, supporting the weak inhibition

and high IC_{50} in the presence of Tyr. Mutation to remove His29 may knock out the Tyr binding ability of the protein.

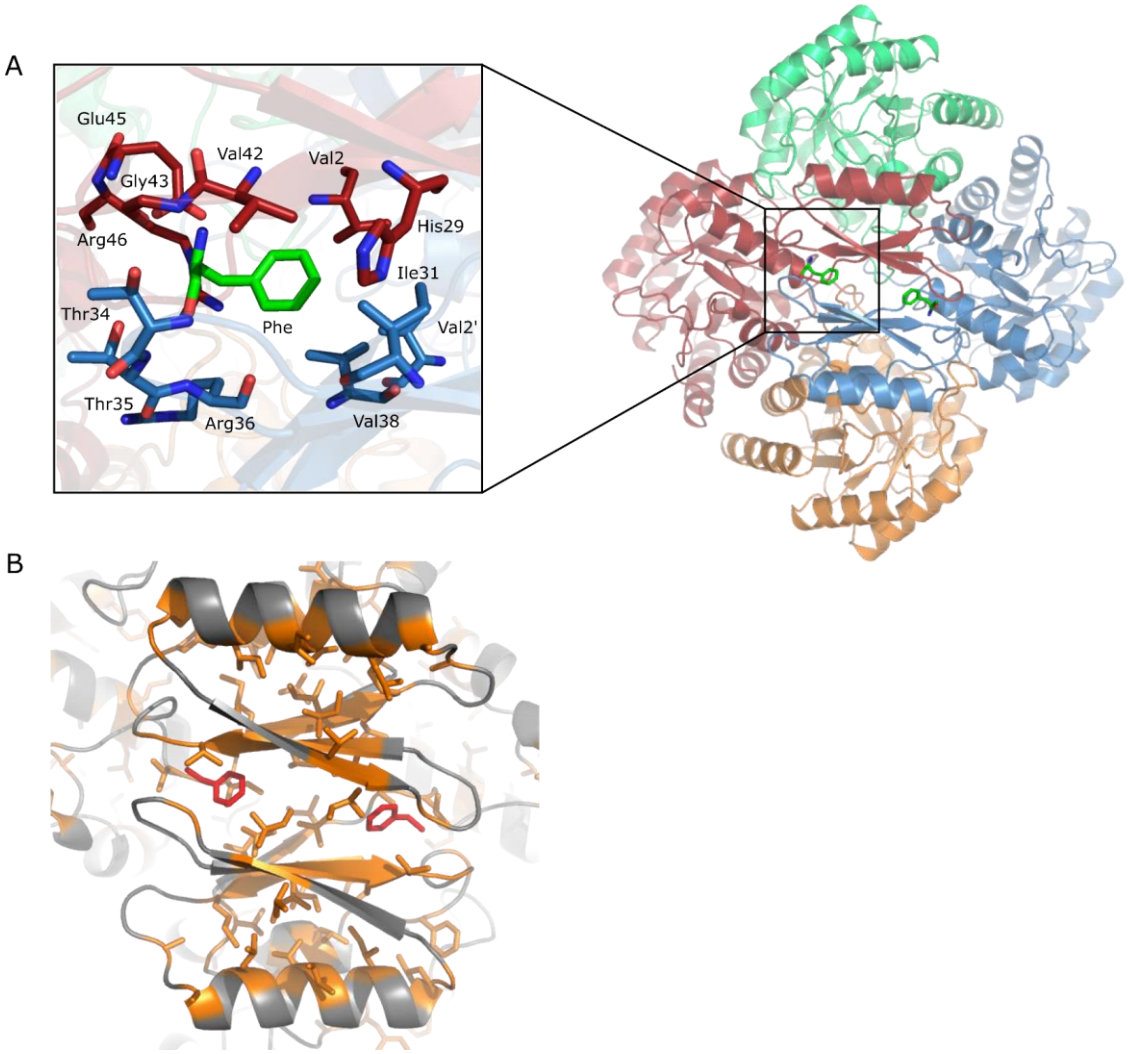


Figure 2.21. A. Phe binding pocket of *Tye*DAH7PS as predicted based on homology model. Residues that contribute to stabilisation of Phe are shown as sticks. Phe is shown as green sticks. Four monomers of *Tye*DAH7PS are coloured separately in red, blue, orange and lime. B. Closed-form model of *Tye*DAH7PS with side-chains of hydrophobic residues highlighted in orange and Phe highlighted in red.

2.6.2. Ligand specificity

The allosteric regulation of *Tma*DAH7PS and *Tye*DAH7PS are both conferred by discrete ACT domains. As introduced in Chapter 1.4, the ACT domain is a common ligand binding module believed to occur early in evolution, and is generally combined with other protein folds.[98, 131] Unlike other DAH7PSs in which allosteric binding sites are preformed, oligomerisation of ACT domains contributes to the construction of ligand binding sites at the domain interfaces.[98] Despite the similar structural fold of ACT domains, their low sequence identity is thought to be associated with the ability to bind a broad range of ligands.[98] This diverse ligand binding ability of ACT domains is of substantial interest in understanding the molecular determinants of ligand specificity in ACT domain containing proteins.

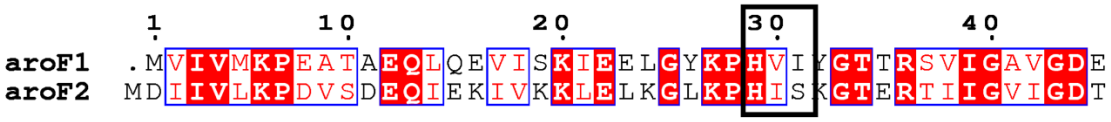


Figure 2.22. Partial sequence alignment of two Type Iβ *Tye*DAH7PSs. The key ligand binding motif in ACT domain is highlighted in black rectangle.

The wild-type *Tma*DAH7PS displays strong inhibition by Tyr and a reduced level of inhibition by Phe. Only one protein homologous to a DAH7PS can be identified in the *T. maritima* genome, indicating that the level of Tyr is sufficient in reporting cellular requirements for the DAH7PS activity and feedback regulates entry to the shikimate pathway. In contrast, there are two proteins homologous to a DAH7PS in the *T. yellowstonii* genome, encoded by *aroF1* (examined in this study) and *aroF2*. Both proteins appear to be Type Iβ DAH7PSs, sharing 55.4 % sequence identity and 73% similarity (Appendix B, Figure S4). Interestingly, instead of having the 'His-X-Ile' motif at the ligand binding site, the *aroF2* DAH7PS has 'His-X-Ser'

at the equivalent site (Figure 2.22). This suggests that the two Type Iβ isozymes in the organism are each responsible for reporting the cellular level of Phe (*aroF1*) or Tyr (*aroF2*).

The Ser or Ile mutation of *Tma*DAH7PS and *Tye*DAH7PS demonstrates the important role of the residue at position 31 of the ACT domain. The location of this residue is in the middle of the interface formed between the β-sheets of the two ACT domains, where the side-chain of ligand Phe or Tyr binds to the pocket (Figure 2.23). Sensibly, the electrostatic character of this position determines the ligand selectivity. If Ser is present, its hydroxyl side-chain forms key hydrogen bonding interaction with Tyr, allowing Tyr binding and subsequent inhibition to occur. Although His29 ensures correct positioning of Ser31, by itself it is not sufficient to stabilise Tyr. If Ile is present at equivalent position, Tyr binding cannot be stabilised and the additional hydrophobicity allows binding of Phe. Single amino acid substitution between Ser and Ile at residue 31 is sufficient to interchange ligand binding specificity between Tyr and Phe.

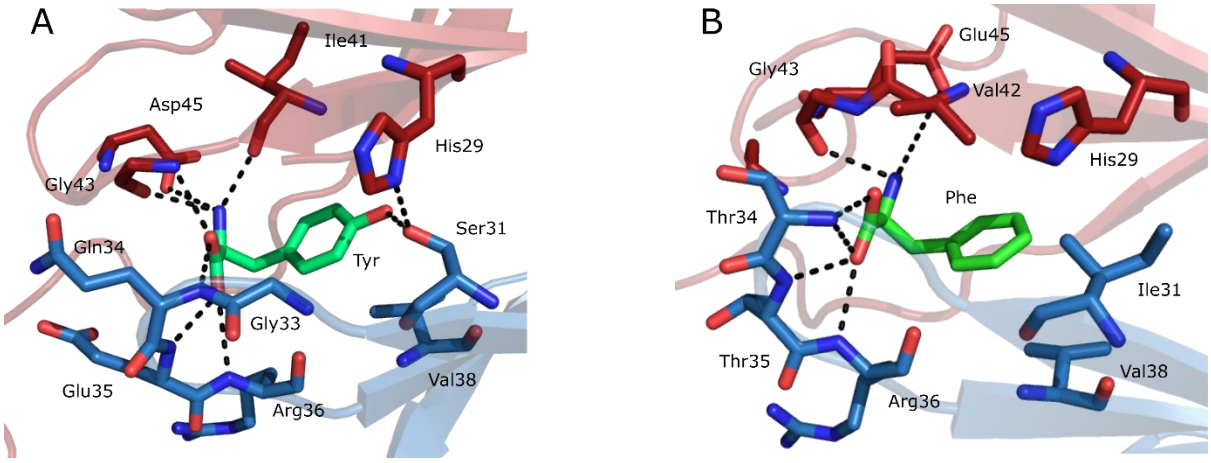


Figure 2.23. Ligand binding at ACT domains of *Tma*DAH7PS (A) and *Tye*DAH7PS (B). A. Tyr binding to *Tma*DAH7PS (PDB 3PG9). B. Phe binding to *Tye*DAH7PS (homology model). Tyr and Phe are shown as green sticks.

2.6.3. Oligomeric state

Despite the highly similar sequences, structures and functions of *Tye*DAH7PS and *Tma*DAH7PS, one of the most significant differences between the two Iβ DAH7PSs is their oligomeric state. *Tma*DAH7PS is solely tetrameric, whereas *Tye*DAH7PS is in a dimer-tetramer equilibrium. This experimentally identified variation is supported by analysis of the tetramer interfaces of the two proteins. The interfaces in *Tye*DAH7PS model was compared with those in the crystal structure of *Tma*DAH7PS (PDB 1RZM) using program PISA.[132] Slightly smaller interface area was predicted in the tetramer interface of *Tye*DAH7PS (average 1377 $Å^2$) compare to *Tma*DAH7PS (1417 $Å^2$). *Tye*DAH7PS also lacked multiple hydrogen bonding and salt bridge interactions, such as at Arg192 (*Tma*DAH7PS residue numbering) at the tetramer interface, which supports its dimer-tetramer equilibrium (Figure 2.24).

Inhibitor binding may promote tetramer formation by dimerisation of the ACT domains, as no repulsive interactions are identified at the tetramer interface. In fact, ACT domain dimerisation is likely a key contributing factor for tetramer formation. Purely tetrameric *Tye*DAH7PS was observed in an N-terminal His_6-tagged version of the protein, evident from SEC-purification and gel filtration experiments where a single peak corresponding to the size of tetrameric species was eluted (Appendix B, Figure S5). This protein was created in reference to a His_6-tagged *Tma*DAH7PS crystal structure (PDB 1VR6), in which His residues from the elongated N-terminus interact with the opposing ACT domains and inadvertently lock the protein in the 'closed-like' inactive conformation without Tyr present (Figure 2.25). The N-terminal His_6-tagged *Tye*DAH7PS may resemble this crystal structure of *Tma*DAH7PS and exists as 'closed tetramers', as no dimeric species was observed from gel filtration experiments. Although further studies are required to validate the predicted 'closed-like' conformation, this construct

may be useful in future crystallisation trials as the complications arisen from the oligomeric equilibrium are eliminated.

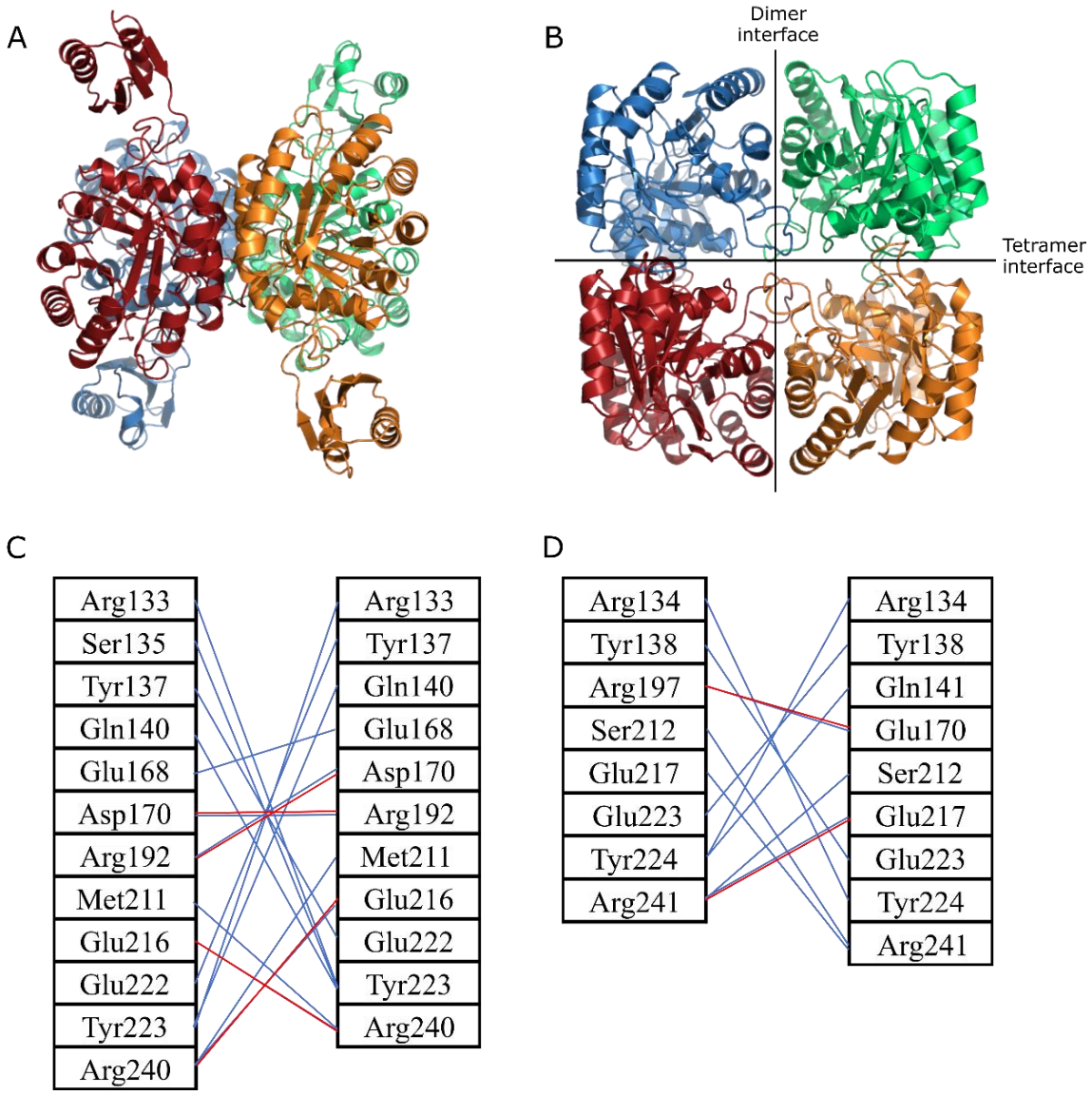


Figure 2. 24. Comparison between tetramer interfaces of *Tma*DAH7PS and *Tye*DAH7PS. A. Side-view of *Tma*DAH7PS with the tight dimer positioned at the front as shown by red and orange, the other pair of tight dimer is shown as blue and green. B. Top-view of *Tma*DAH7PS showing tetramer interface and tight dimer interface. C. Hydrogen bonding (blue) and salt bridge (red) interactions identified at the tetramer interface of *Tma*DAH7PS. D. Hydrogen bonding interactions (blue) and salt bridges (red) predicted at the tetramer interface of *Tye*DAH7PS based on homology model.

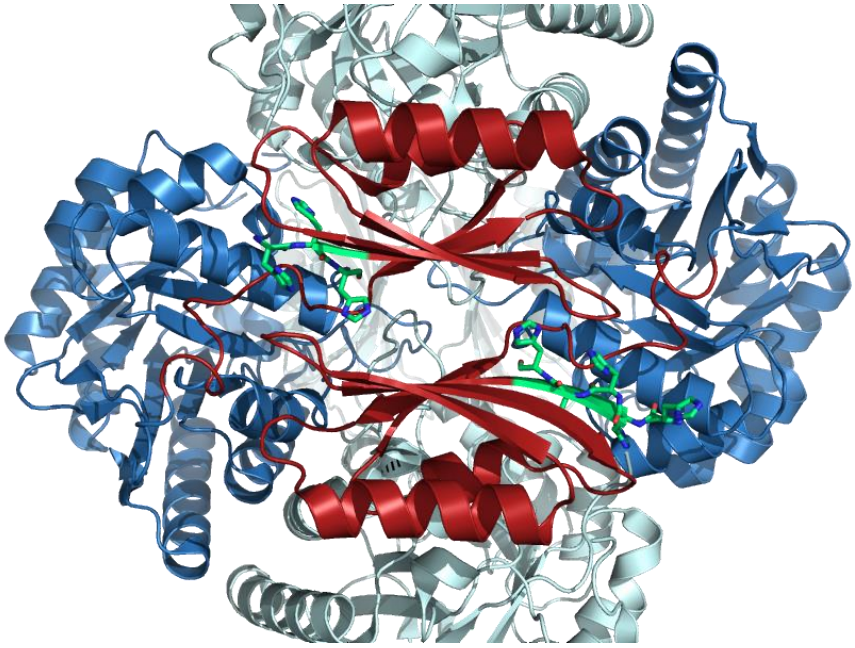


Figure 2.25. N-terminal His tag locks *Tma*DAH7PS in a closed inactive conformation in the absence of allosteric inhibitor (PDB 1VR6). ACT domain is shown in red. DAH7PS barrel is in blue or light blue. His tag is shown as green sticks.

The assembly of oligomeric proteins plays an important role in protein function. For instance, active sites or regulatory sites of some protein are constructed at oligomer interfaces; and oligomerisation also contributes to protein dynamics and stability.[133-138] In general, it appears that, although most characterised Type Iβ DAH7PSs are tetrameric, the tetramer interfaces are weak in many cases. For instance, removal of the ACT domain from *Tma*DAH7PS and deletion of the CM domain from *Gsp*DAH7PS both result in the presence of dimeric species, while the wild-type proteins are solely tetrameric.[63, 78] Additionally, the unregulated *Pfu*DAH7PS becomes dimeric by a single amino acid substitution at the tetramer interface.[134] Therefore, perhaps unsurprisingly, the newly characterised member of Type Iβ DAH7PS from *T. yellowstonii* is naturally in a dimer-tetramer equilibrium. Future experiments testing the impact of factors such as salt concentrations, temperature, or pH on the equilibrium may help further the understanding of this equilibrium. Truncation of the ACT domain from *Tye*DAH7PS may

also aid in defining more clear the role of the ACT domain in delivering the oligomeric state of the protein.

The tetrameric form of *Tye*DAH7PS appears to be responsible for allosteric regulation as the addition of Phe promotes formation of tetramer from dimer, perhaps largely due to dimerisation of the ACT domains. Consistent with the native TIM barrels being functional as a dimer,[135] it is possible that the *Tye*DAH7PS dimer is primarily responsible for providing DAH7PS activity in the organism, and the tetramer is responsible for providing allosteric regulation as the protein can only be inhibited when tetramer is formed, although the role of oligomerisation in protein stability is not yet studied for *Tye*DAH7PS. The existence of functional dimer appears to be a recurring theme and has been reported in Type Iβ, Type Iα and Type II DAH7PSs, for which further oligomerisation delivers protein stability and/or allostery.[134, 139-140] However, it is important to note that whether the dimeric species of *Tye*DAH7PS occur *in vivo* is not yet known, and the biological function of the dimer (if any) is yet to be determined. The cellular conditions may also alter the oligomeric equilibrium.

Chapter 3. Interchangeability of modular allostery

Preface

Most proteins comprise two or more domains from a limited suite of protein families. These domains are often rearranged in various combinations through gene fusion events to evolve new protein functions, including the acquisition of protein allostery through the incorporation of regulatory domains. The enzyme DAH7PS is the first enzyme of aromatic amino acid biosynthesis and displays a diverse range of allosteric mechanisms as described in previous chapters. DAH7PSs adopt a common architecture with a shared $(\beta/\alpha)_8$ catalytic domain which can be attached to an ACT-like or a chorismate mutase regulatory domain that operate via distinct mechanisms. These domains confer allosteric regulation by controlling DAH7PS function in response to ligand Tyr or prephenate, respectively. Starting with contemporary DAH7PS proteins, two protein chimera were created, with interchanged regulatory domains. Both engineered proteins are catalytically active, and deliver new functional allostery with switched ligand specificity and mechanisms provided by their regulatory domains. This interchangeability of protein domains represents an efficient method to not only engineer allostery in multidomain proteins but to create a new bifunctional enzyme. Part of the work described in this chapter is adapted from a manuscript submitted for publication.

3.1. Introduction

As briefly described in Chapter 1, protein allostery is central to the regulation of many biological processes, ranging from ligand transport to metabolic function and control. Allostery occurs when ligand binding at one site communicates with a remote functional site, resulting in a change in protein function. Allosteric regulation of protein functions often involves a complex network of interactions to deliver signal between distal sites. Signal communication is achieved via diverse mechanisms ranging from large conformational changes to subtle changes in protein dynamics. [1, 141-142] The understanding of these remote communications is of great interest, particularly in the fields of drug design and protein engineering.[143-144]

Most proteins contain two or more domains.[145] These domains and their interactions govern the function of a protein and are considered evolutionary units for modular assembly of new protein architectures.[146-148] Biological data suggest that only a limited number of protein folds exists in nature, and protein functions evolve from mutation, duplication and recombination of ancestral genes under selective pressure.[149] Domain recombination via gene fusion events represents one of the major pathways for the evolution of allostery.[150-153] This principle has been recognised in many natural systems.[151]

The enzyme DAH7PS, situated at a crucial biosynthetic node in the shikimate pathway, is often precisely feedback-regulated to control pathway flux in response to metabolic demand via a variety of allosteric strategies in different organisms (Chapter 1.3). Most characterised DAH7PS enzymes are tetrameric in solution, with each chain comprising a $(\beta/\alpha)_8$ catalytic barrel. This barrel is frequently decorated with additional structural elements, which are responsible for conferring allostery.[80] One major class of DAH7PS enzymes (Type Iβ) share a

catalytic domain of a structurally uninterrupted classic TIM barrel $(\beta/\alpha)_8$ fold, and are either unregulated or allosterically regulated (Figure 3.1 A). The simplest form is composed solely of the barrel without any additional domain, such as *Pfu*DAH7PS, and hence is unregulated.[72-73, 134] The regulated DAH7PSs in this group display discrete domains appended to the catalytic barrel at N- or C- terminus. These domains are either an ACT-like domain[98] or a CM (AroQ) domain, which possesses CM enzymatic activity. The attached ACT-like and CM domains deliver allostery on a tetrameric DAH7PS scaffold via physical gating of the active site associated with the binding of an allosteric ligand (Figure 3.1 B, C).[63, 77-78] This is best exemplified in the well-characterised DAH7PS from *Thermotoga maritima*, which undergoes a remarkable conformational change in response to the presence of Tyr, in which the ACT-like domains from opposing chains dimerise to form binding sites for the allosteric ligand and block substrate access (Chapter 2).[63] This enzyme is also partially inhibited by Phe, although to a reduced extent. Similarly, prephenate binding to the CM domain in DAH7PS from *Geobacillus sp* is associated with the more intimate association between the CM and DAH7PS domains to form a compact structure and limit catalysis (Figure 3.1 C).[78]

It has been demonstrated that the allosteric domain of the *T. maritima* DAH7PS can be transferred onto the unadorned, unregulated DAH7PS of *P. furiosus* to confer an allosteric response in the latter DAH7PS.[117] We propose that the recruitment of a regulatory domain is a general strategy that by itself is sufficient for providing allosteric control of enzymes. To validate this hypothesis and to demonstrate that the key information associated with delivery of allostery resides in the regulatory domain, this chapter explores the interchangeability of the two distinct regulatory strategies employed by different allosterically controlled DAH7PS enzymes, through construction of protein variants that mix and match catalytic and regulatory domains of DAH7PSs from *T. maritima* and *G. sp*. These studies illustrate the remarkable ease with which functional allostery can be acquired by gene fusion events, and provide insights

into the evolution of modular allostery, in which existing ligand-binding domains or enzymes can be repurposed to provide allostery.

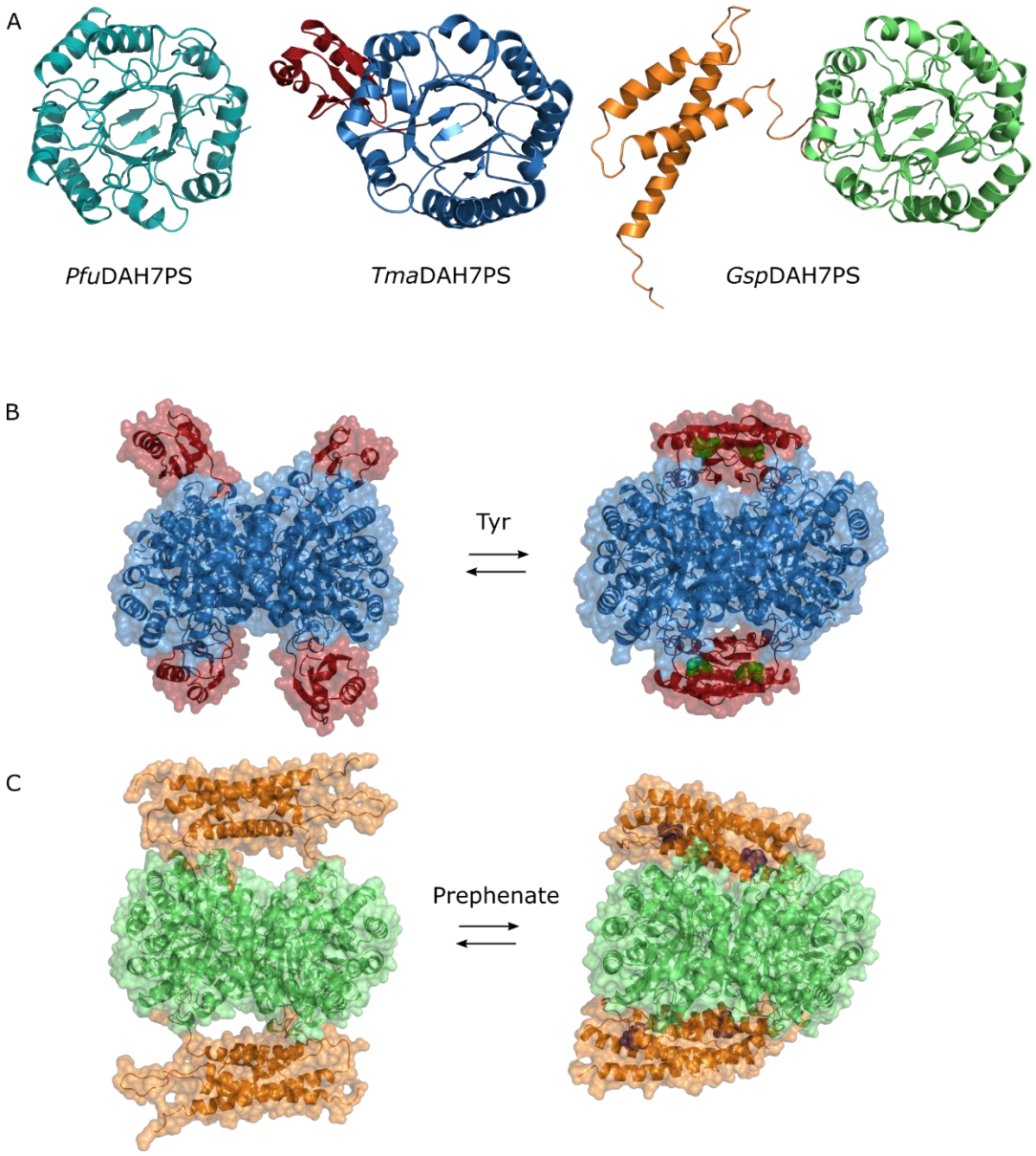


Figure 3.1. Structure and allostery of Type Iβ DAH7PS. A. Monomeric unit of DAH7PS from *P. furiosus*, *T. maritima*, and *G. sp* share an uninterrupted catalytic barrel. ACT domain in *Tma*DAH7PS is shown in red; and CM domain in *Gsp*DAH7PS is shown in orange. B. Both *Tma*DAH7PS and *Gsp*DAH7PS display significant conformational changes in the presence of ligand, compared to the apo forms (Tyr is shown as green spheres, and prephenate is shown as navy spheres).

3.2. Design and preparation of protein variants

Crystal structures of the *Tma*DAH7PS and *Gsp*DAH7PS suggest similar architecture and arrangement of the homotetramers with the catalytic barrels connected to their respective N-terminal regulatory domains via a β-hairpin and a flexible linker region. The linker region is crucial for the interaction between regulatory and catalytic domains in both enzymes and allowing appropriate conformational change between active and inhibited states.[63] The amino-acid regions that incorporate regulatory domains, linker regions and catalytic domains of *Tma*DAH7PS and *Gsp*DAH7PS were identified individually from sequence and structure alignments (Figure 3.2). Two protein variants with exchanged regulatory and catalytic domains were designed: ^{Tma}ACT-^{Gsp}DAH7PS with the regulatory domain and linker region from *Tma*DAH7PS (residue 1-93) and the catalytic domain from *Gsp*DAH7PS (residues 118-362), and the complementary chimera ^{Gsp}CM-^{Tma}DAH7PS with the regulatory domain and the linker region from *Gsp*DAH7PS (residues 1-117) and catalytic domain from *Tma*DAH7PS (residue 94-338). Both constructs were created by amplifying each segment from the corresponding parent wild-type gene with overlap and fusing the amplified products (Table 7.1). The fused gene fragments encoding each chimeric protein were cloned using Gateway© technique.

^{Tma}ACT-^{Gsp}DAH7PS was transformed in pDEST14 and expressed in BL21(DE3)* cells, and its expression and purification were performed following previously published procedures for *Tma*DAH7PS (Appendix C, Figure S6).[63] Briefly, expression was induced with 0.5 mM IPTG followed by overnight growth at 23 °C. Two-step purification involved hydrophobic interaction chromatography (SourcePhe) and size-exclusion chromatography (Superdex 200 26/60). The ^{Gsp}CM-^{Tma}DAH7PS was transformed in pDEST15 and expressed in BL21(DE3)pLysS cells using the same expression conditions, and purified using GSTrap HP column following published procedures for the truncated *Gsp*DAH7PS.[78] Size-exclusion chromatography was

also used for final step of purification in buffer containing 10 mM BTP (pH 7.5), 100 mM KCl, and 200 µM PEP.

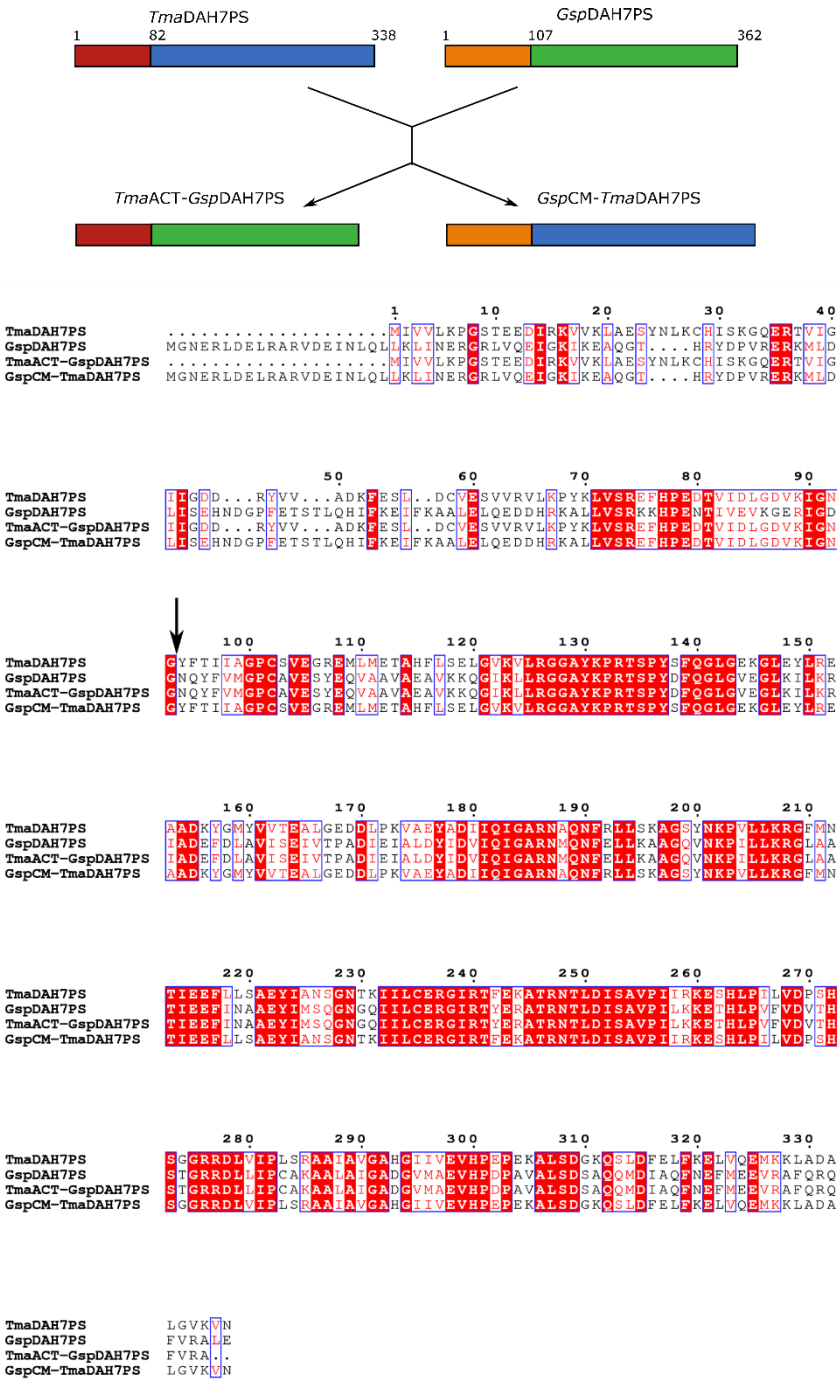


Figure 3.2. Scheme of chimera design. Red represents ACT domain from *Tma*DAH7PS, orange represents CM domain from *Gsp*DAH7PS, blue represents catalytic domain of *Tma*DAH7PS domain,

and green represents catalytic domain of *Gsp*DAH7PS. Alignment includes sequences of the wild-type parent proteins and two protein chimera. Black arrow identifies the fusion site.

3.3. Results

3.3.1. Function of the chimeric proteins

The wild-type parent proteins *Tma*DAH7PS and *Gsp*DAH7PS share similar homotetrameric quaternary structures and tertiary structures with each chain composed of an N-terminal regulatory domain (respectively, ACT or CM) attached to a catalytic $(\beta\alpha)_8$-barrel housing the active site. The *Tma*DAH7PS and *Gsp*DAH7PS catalytic barrels share moderate sequence identity of 56% (Figure 3.2). Two protein variants were created with exchanged regulatory and catalytic domains based on the analysis of wild-type sequences and architectures. Kinetic parameters for DAH7PS and/or CM activities of the chimeric proteins were determined following previously described procedures by measuring consumption of chorismate at 274 nm or PEP at 232 nm, with each reaction containing 0.05 μM enzyme.[78] Kinetic parameters K_M and k_{cat} were determined by fitting data into the Michaelis–Menten equation by using GraphPad Prism 7 software.

Both domain swapped variants delivered DAH7PS catalysis, albeit with some alterations in their catalytic efficiencies (Table 3.1). Compared to the wild-type *Tma*DAH7PS, ^{Gsp}CM-^{Tma}DAH7PS, which shares the same catalytic core, exhibited impaired activity with significantly decreased k_{cat}/K_M values for both PEP and E4P substrates, implying there may be some restriction of access for substrates to the catalytic centre introduced by the fused CM domains. The origin of this may lie in the fact that the CM domain is dimeric both in the absence and presence of the allosteric ligand, whereas the ACT-like domain only dimerises upon ligand

binding.[63, 78] Hence, perhaps unsurprisingly, the adoption of the more restricted *Gsp*CM domain to the *Tma*DAH7PS catalytic core was accompanied by attenuation of catalysis, whereas ^{Tma}ACT-^{Gsp}DAH7PS displayed an approximately two-fold boost in catalytic efficiency compared to the wild-type *Gsp*DAH7PS. That physical constraints restrict catalysis by the DAH7PS core barrel is supported by the higher activity displayed by truncated forms of *Tma*DAH7PS and *Gsp*DAH7PS in which, respectively, the ACT and CM domains have been removed.[63, 78]

Table 3.1. Kinetic parameters for the two protein variants and their wild-type counterparts.*

	DAH7PS activity					CM activity		
	K_M^{PEP} (μM)	K_M^{E4P} (μM)	k_{cat} (s^{-1})	k_{cat}/K_M^{PEP} ($\mu M^{-1}s^{-1}$)	k_{cat}/K_M^{E4P} ($\mu M^{-1}s^{-1}$)	$K_M^{chorismate}$ (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)
^{Gsp}CM-^{Tma}DAH7PS	34 ± 2.7	19 ± 1.8	4.8 ± 0.1	0.14	0.25	190 ± 16	4.1 ± 0.1	0.02
^{Tma}ACT-^{Gsp}DAH7PS	57 ± 1.6	54 ± 1.4	60 ±2	1.05	1.11	NA	NA	NA
***Gsp*DAH7PS^{WT}**	45 ± 4	61 ± 8	28 ± 0.9	0.62	0.46	98 ± 7	3.4 ± 0.1	0.03
***Tma*DAH7PS^{WT}**	8.4 ± 0.7	15 ± 1	14 ± 0.3	1.67	0.93	NA	NA	NA

*Concentration of E4P was varied between 5 μM and 400 μM while PEP was held at 295 μM. Concentration of PEP was varied between 6 μM and 350 μM while E4P was fixed at 310 μM. Reactions contained enzyme, PEP, and 100 μM Mn^{2+} (for proteins containing catalytic barrel from *T. maritima*) or Cd^{2+} (for proteins containing catalytic barrel from *G. sp.*), and were equilibrated in 50 mM BTP buffer at 60 °C (pH 7.5) before E4P was added to initiate the reaction.

Unlike the ACT domain, which has only a ligand-binding role, the regulatory domain of full-length *Gsp*DAH7PS has both catalytic and regulatory functions.[78] The CM catalytic activity was comparably maintained when the domain was transferred to the alternative catalytic barrel in ^{Gsp}CM-^{Tma}DAH7PS (Table 3.1), although this chimera displayed an approximately two-fold

increase in K_M for chorismate compared to wild-type *Gsp*DAH7PS. As expected, with no CM domain, the variant ^{Tma}ACT-^{Gsp}DAH7PS displayed only DAH7PS activity.

As with all wild-type DAH7PS enzymes characterised to date, the catalytic activity of the protein chimera depends on presence of a divalent metal ion (Figure 3.3 A). Mn^{2+} is most activating metal ion for the wild-type *Tma*DAH7PS, whereas Cd^{2+} delivers maximal activity for *Gsp*DAH7PS.[63, 78] A range of metal ions were tested with the protein chimera. Largely as expected, each of the DAH7PS catalytic cores reflected their inherent metal preference, with ^{Tma}ACT-^{Gsp}DAH7PS highly favoring Cd^{2+}, whereas ^{Gsp}CM-^{Tma}DAH7PS showed more than 90% activity in the presence of Mn^{2+} or Cd^{2+}.

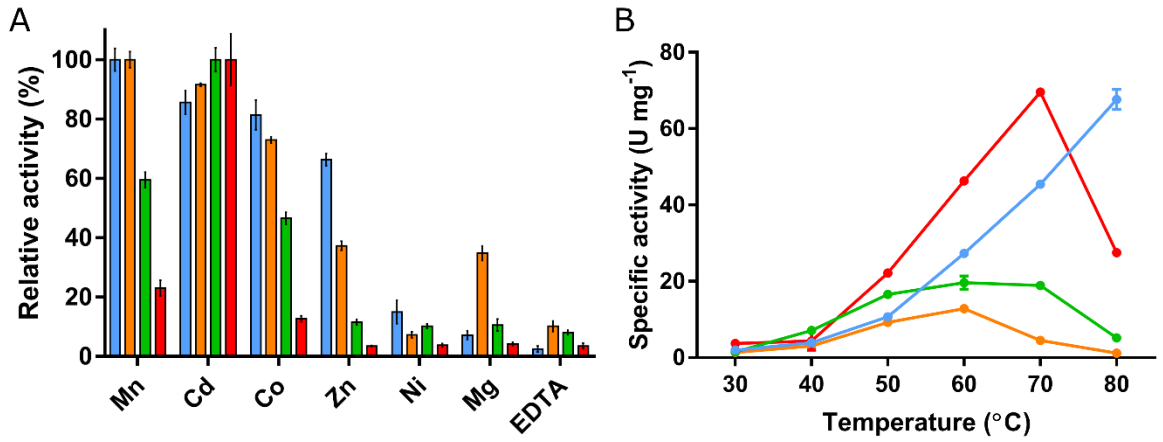


Figure 3.3. Catalytic activity of the chimera and parent proteins. A. DAH7PS activities of *Tma*DAH7PS^{WT} (blue), *Gsp*DAH7PS^{WT}(green), ^{Tma}ACT-^{Gsp}DAH7PS (red) and ^{Gsp}CM-^{Tma}DAH7PS (orange), in the presence (or absence) of a range of metals. B. The effect of temperature on specific activity of the four enzymes in the same colour coding as Figure 3.3 A. Triplicate measurements were performed and SD values were used as errors.

Activity profiles of the two chimeric proteins and parent proteins with elevated temperatures (30 °C-80 °C) were also assessed using the standard methods for kinetic measurements of

DAH7PS activity (Chapter 7). Consistent with the thermophilic properties of the wild-type enzymes, both enzymes became more active at elevated temperatures, reaching their optimal activity at temperatures above 60 °C (Figure 3.3 B). Intriguingly, activity of ^{Tma}ACT-^{Gsp}DAH7PS was enhanced significantly at elevated temperatures compared to the wild-type *Gsp*DAH7PS. On the other hand, ^{Gsp}CM-^{Tma}DAH7PS appeared to be least active across all temperatures. Proteins with the *Gsp*CM domain, chimeric and wild-type, generally displayed lower activities than proteins with the *Tma*ACT domain at the temperatures tested. This change in activity profile may relate to the structural difference in the regulatory domains and the different optimal temperatures of the parent proteins.

3.3.2. Allosteric regulation

To test the effect of potential allosteric ligands, inhibition assays were performed for both DAH7PS chimera and compared with the response of the wild-type proteins (Figure 3.4, Appendix C, Figure S7). Consistent with the presence of a regulatory domain in both chimera, inhibition of both enzymes was observed. Presenting the ACT domain, the DAH7PS activity of ^{Tma}ACT-^{Gsp}DAH7PS was reduced by 70% at high Tyr concentrations, and approximately 30% at high Phe concentrations. This difference in ligand sensitivity for Tyr and Phe was comparable with that of the wild-type *Tma*DAH7PS, for which Tyr exhibited more significant inhibitory effect than Phe, although the IC_{50} value of Tyr for the chimera (210 µM) was almost 10-fold higher than that of the wild-type enzyme (22.5 µM). Presence of the CM domain in ^{Gsp}CM-^{Tma}DAH7PS delivered sensitivity towards prephenate, although the inhibition was less profound than that for the wild-type *Gsp*DAH7PS. The presence of 77 µM prephenate reduced DAH7PS activity to 50% of its uninhibited value, compared to 20 µM required for the same attenuation of wild-type protein catalytic activity. The maximum level of inhibition was also

altered; the chimeric protein retained 31% activity at higher prephenate concentrations in comparison to the 4% residual activity observed for the wild-type enzyme.

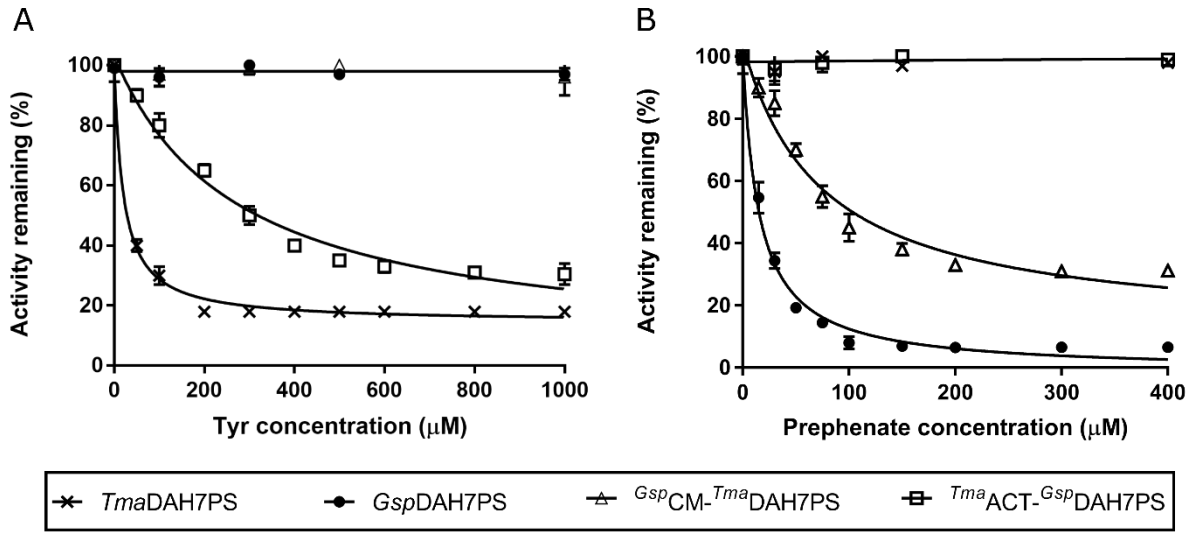


Figure 3. 4. Inhibition of DAH7PS activity in parent proteins and chimera. A. Wild-type *Tma*DAH7PS and ^{Tma}ACT-^{Gsp}DAH7PS showed sensitivity towards Tyr. B. Wild-type *Gsp*DAH7PS and ^{Gsp}CM-^{Tma}DAH7PS displayed impaired catalysis with increasing prephenate concentrations.

Binding of allosteric ligands stabilised the parent proteins and the chimera as demonstrated by differential scanning fluorimetry (DSF) results (Figure 3.5, Appendix C, Figure S8). Experimental details of the thermostability measurements are included in Chapter 7. In the presence of inhibitors, the melting temperatures of both chimera increased by around 3 °C, in close agreement with the degree of stabilisation for the wild-type proteins. Interestingly, the thermostability of ^{Tma}ACT-^{Gsp}DAH7PS improved almost 20 °C compared to the wild-type *Gsp*DAH7PS. On the other hand, the stability of the ^{Gsp}CM-^{Tma}DAH7PS appeared to be largely determined by the *Tma*DAH7PS core with the denaturing event occurring at above 96 °C. The thermostability profiles exhibited by the chimera largely reflect the difference in inherent thermostability of the wild-type proteins, remarkably, even when the transferred domain is only a fraction (~1/3) of the core DAH7PS barrel, as in ^{Tma}ACT-^{Gsp}DAH7PS.

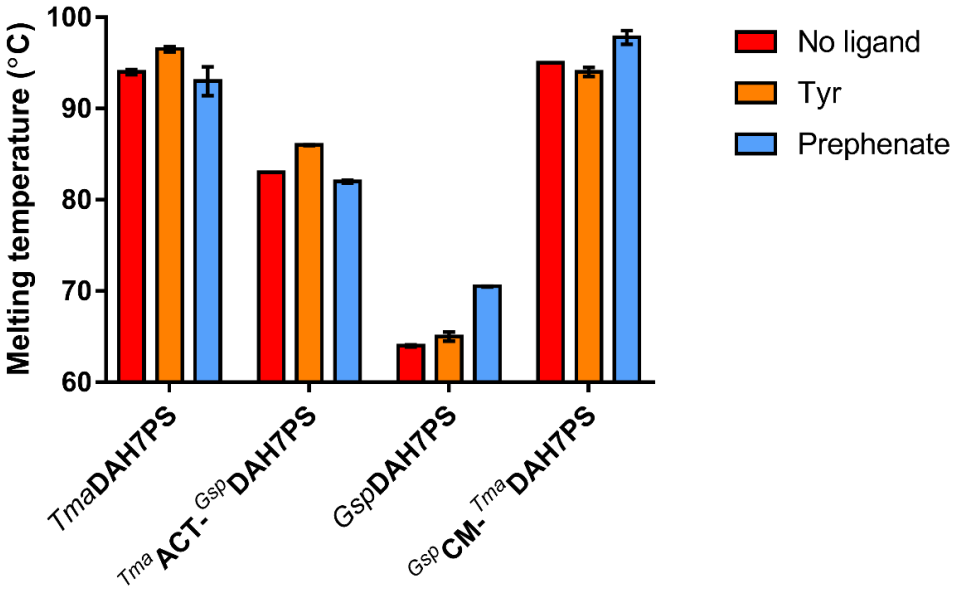


Figure 3.5. Thermostability of the parent and chimeric proteins in the presence or absence of ligand.

3.3.3. Conformational change

The allosteric mechanisms of both parental proteins involve conformational rearrangements as observed in solution by SAXS experiments.[63, 78] To investigate the allosteric strategies employed by the chimera, SAXS experiments were performed. In the presence of Tyr, *Tma*ACT-*Gsp*DAH7PS became more compact with a reduced *R*g value (derived from Guinier plot) of 32.8 ± 0.2 Å, compared to the *R*g of the apo form 34.0 ± 0.2 Å (Figure 3.6 A, Table 3.2). The Kratky plot displayed a more defined curve in the presence of Tyr than without Tyr, consistent with the decreased flexibility of the protein (Figure 3.6 B). The *Gsp*CM-*Tma*DAH7PS also showed a conformational change with addition of prephenate, with *R*g decreasing from 36.9 ± 0.2 Å to 34.8 ± 0.2 Å, consistent also with the more defined Kratky plot (Figure 3.6 C and D, Table 3.2).

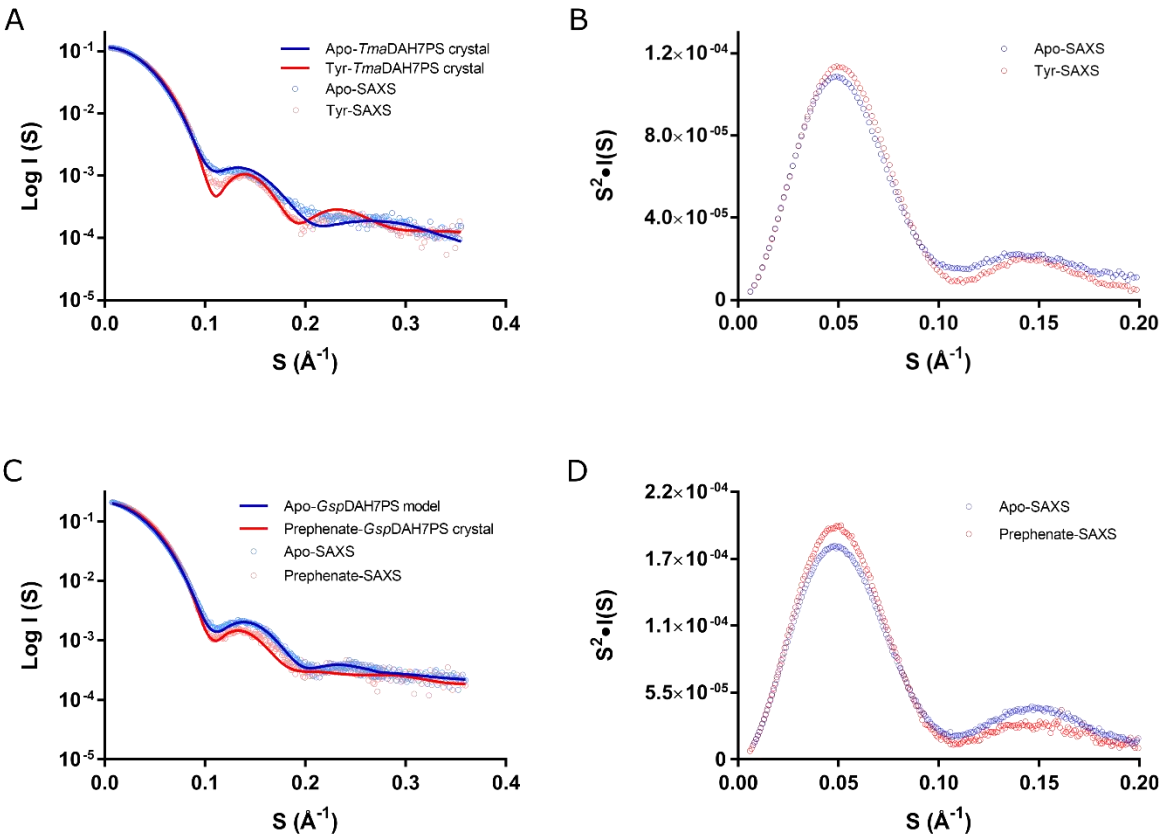


Figure 3.6. SAXS profiles of the chimera. A. ^{Tma}ACT-^{Gsp}DAH7PS SAXS scattering profiles fitting with the appropriate theoretical scatterings calculated from *Tma*DAH7PS crystal structures. B. Kratky plot of ^{Tma}ACT-^{Gsp}DAH7PS SAXS profiles in the absence or presence of Tyr. C. ^{Gsp}CM-^{Tma}DAH7PS SAXS scattering profiles fitting with the appropriate theoretical scatterings calculated from *Gsp*DAH7PS crystal structure (closed form) or model (open form). D. Kratky plot of ^{Gsp}CM-^{Tma}DAH7PS SAXS profiles in the absence or presence of prephenate.

To compare these conformational changes observed in the protein chimera, with those observed for the wild-type proteins, the SAXS profiles were fitted with the calculated theoretical scattering from crystal structures or homology model of each parent protein (Figure 3.6, Table 3.2). Scattering profiles for ^{Tma}ACT-^{Gsp}DAH7PS showed a good fit with the open-form of the *Tma*DAH7PS crystal structure, but not the closed form, in the absence of Tyr. Conversely, in

the presence of Tyr the scattering profile of ^{Tma}ACT-^{Gsp}DAH7PS gave a better fit with the closed, Tyr bound *Tma*DAH7PS than with the open apo form. In the absence of prephenate, the scattering profile of ^{Gsp}CM-^{Tma}DAH7PS fitted closely with the open-form model of the wild-type *Gsp*DAH7PS and poorly with the closed, prephenate-bound form of *Gsp*DAH7PS. In the presence of prephenate, ^{Gsp}CM-^{Tma}DAH7PS was better fitted by the closed prephenate-bound form of *Gsp*DAH7PS than by the open apo form.

Table 3.2. Structural parameters from SAXS profile of each chimera in the absence or presence of Tyr or prephenate.

Structural parameters	Apo-^{Tma}ACT-^{Gsp}DAH7PS	Tyr-^{Tma}ACT-^{Gsp}DAH7PS	Apo-^{Gsp}CM-^{Tma}DAH7PS	Prephenate-^{Gsp}CM-^{Tma}DAH7PS
I(0) (cm^{-1}) [from P(r)]	0.12 ± 0.003	0.116 ± 0.002	0.408 ± 0.002	0.397 ± 0.003
R_g (Å) [from P(r)]	33.96 ± 0.08	32.71 ± 0.08	36.89 ± 0.2	34.76 ± 0.2
I(0) (cm^{-1}) (from Guinier)	0.12 ± 0.001	0.116 ± 0.001	0.408 ± 0.001	0.397 ± 0.001
D_{max} (Å)	104.16	101.35	120.71	118.43
R_g (Å) (from Guinier)	34.04 ± 0.2	32.75 ± 0.2	36.91 ± 0.2	34.77 ± 0.2
Porod volume estimate ($Å^3$)	226,364	222,091	274,596	258,918
Fit with crystal/model structure (χ^2) (from CRYSOL)	1.08	0.7	0.78	1.33

3.4. Summary and discussion

Design of proteins with new properties and functions is an important goal of biotechnology. The modular feature of many natural proteins suggests that common building domains and modules are likely to have the evolutionary advantages of being autonomous and portable, offering simplicity in recombination to generate new functions by allowing transfer of information through their interactions.[154] Statistical coupling analysis of several naturally fused

protein systems revealed networks of coevolving amino acids involved in allosteric communication, and these networks displayed strong connectivity, proposing the feasibility of engineering artificial allosteric systems by transfer of allosteric networks between proteins.[144, 155] Although this transferability has not been extensively explored, a few studies indicate its plausibility. For example, a modulating maltose binding protein was fused to an unrelated enzyme β-lactamase, introducing maltose regulation to the unregulated β-lactamase.[156-158] In another example, ultrasensitive molecular switches were built by modular recombination of multiple SH3-peptide autoinhibitory interactions on the WASP protein to introduce strong cooperativity with respect to the ligand SH3.[159-160] Also, an *E. coli* dihydrofolate reductase (DHFR) enzyme was coupled with a light-sensing protein from plants to generate a light-sensitive DHFR.[144] These studies demonstrate that methods of 'mix and match' between modular components can be effective ways to create substantial functional changes associated with allostery and expand the repertoire of artificial proteins.

Studies described in this chapter demonstrate for the first time the *interchangeability* of regulatory domains in a homo-oligomeric protein. By simple gene recombination of contemporary sequences, the Tyr-binding ACT domain on one DAH7PS can be interchanged with a prephenate-binding CM domain from another DAH7PS. This interchange functionally swaps allosteric regulation in both DAH7PS enzymes elicited by formation of or changes to a dimeric structure upon binding of the appropriate allosteric effector. What is somewhat surprising is that gene fusion to functionally link these contemporary domains requires no modifications to the core catalytic barrels. Fusion of regulatory domain nevertheless inflicts some catalytic penalty on the barrel, and the degree of compensated activity may be associated with the nature of the extended domains. Allostery mediated by the ACT domain and CM domain represent feedback regulatory strategies on different levels, in which the end-products of the shikimate pathway are regulatory reporters for ACT-DAH7PS functions, whereas the

intermediate product reports cellular demand to control CM-DAH7PS function. The interchangeability of the two feedback regulation strategies may reflect the ease in which appropriate regulation can be obtained by the organism based on its demand.

Despite the ease of functional allostery interchange, there are enabling features of the contemporary DAH7PS scaffold that are important in accommodating the recombination and for maintaining allosteric networks. Core barrel oligomerisation is a prerequisite for allostery that involves dimerisation of the regulatory module.[62-63, 72, 78] The DAH7PS barrel homotetramer adopts an overall conformation that supports the allosteric function, with diagonally opposite chains delivering the regulatory domain on either side of the tetramer plane (Figure 3.1). This tetrameric assembly appears to be a common feature of most Type Iβ DAH7PSs, which is shared by both regulated and unregulated DAH7PS,[63, 78, 118, 134] and a related enzyme 3-deoxy-D-*manno*-2-octulosonate 8-phosphate synthase (KDO8PS).[161] Intriguingly, quaternary structure is delicately balanced in this DAH7PS enzyme class. Removal of the ACT-like domain from *Tma*DAH7PS results in dimerisation and the (unregulated) *Pfu*DAH7PS is rendered dimeric by a single amino acid substitution.[63, 134] Association of dimers into tetramers was probably followed by acquisition of terminal regulatory domains.[72] Thus, it appears that the functional gene fusion to deliver allostery was facilitated serendipitously by the adoption of the appropriate homotetrameric catalytic template.

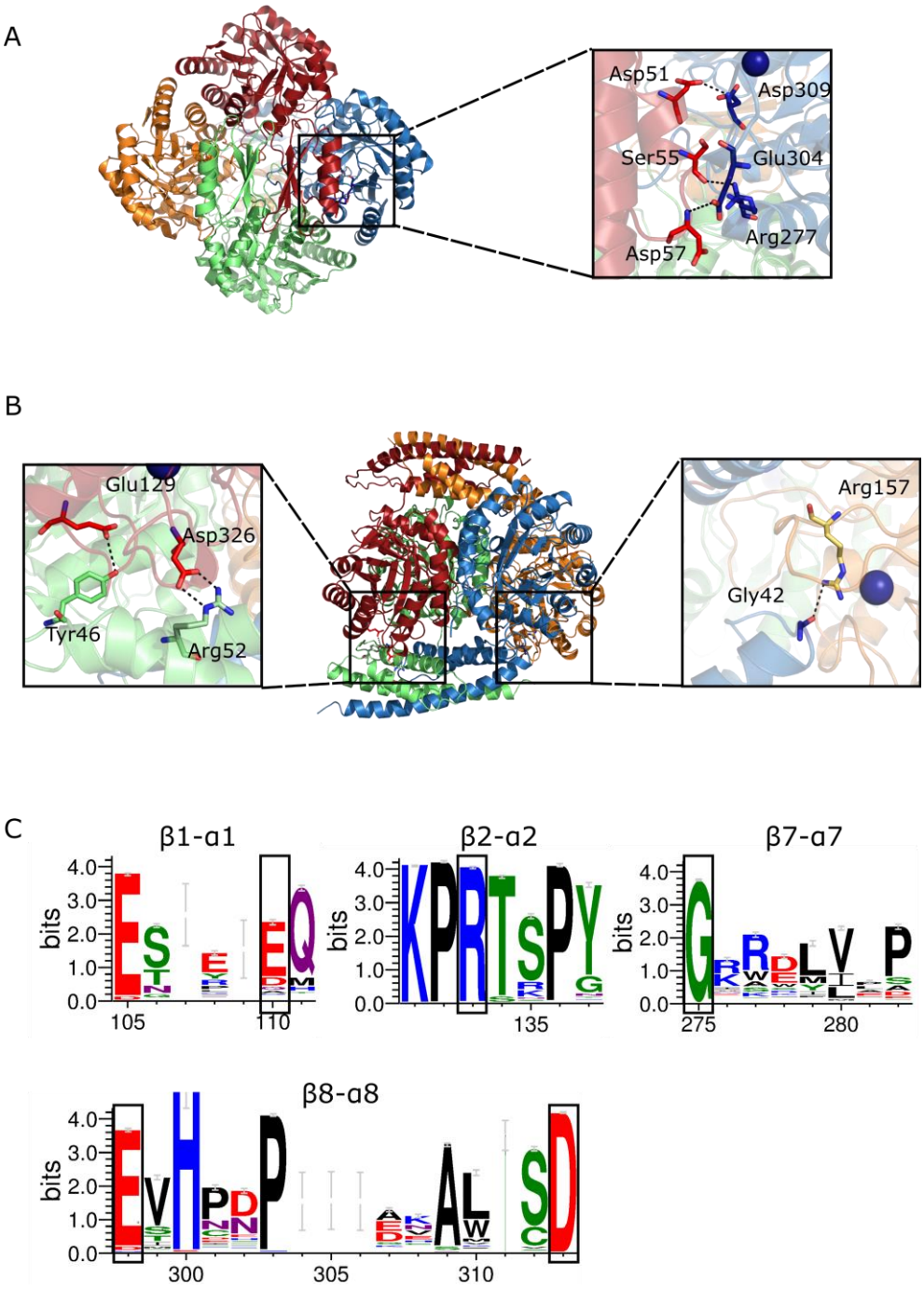


Figure 3.7. Analysis of key interactions associated with allostery. Key residues involved in interactions between regulatory and catalytic domains are identified in *Tma*DAH7PS (top view, A) and *Gsp*DAH7PS (B). C. Partial alignment of Type Iβ DAH7PS sequences, showing segments of the catalytic barrel. Black rectangles highlight the location of key barrel residues identified in Figure 3.7 A and B. Figure was formatted with WebLogo3.

The barrels are not highly conserved overall, but there do appear to be some sequence elements that are associated with delivering allostery in both systems that are shared between *Gsp*DAH7PS and *Tma*DAH7PS. The key hydrogen bond contacts between either ACT domain or the CM domain with the catalytic DAH7PS domain reside mostly at the C-terminus of the diagonally adjacent barrel (Figure 3.7 A, B). In the ligand-bound form of *Tma*DAH7PS, major hydrogen bonds are formed between an ACT domain and the barrel from the adjacent chain, including Asp51 and Asp309, Ser55 and Arg277, Asp57 and Glu304. Due to the asymmetrical nature of the domain arrangement across the vertical plane in the crystal structure *Gsp*DAH7PS, each CM polypeptide chain communicates with its adjacent barrel through different hydrogen bonding interactions. Key interactions are found between the α1-α2 loop of the CM domain with the C-terminal end of the barrel, including Arg52 and Asp326, as well as Tyr46 and Glu29 on one end of the barrel, plus Gly42 and Arg157 on the other end. Sequence alignment between *Tma*DAH7PS and *Gsp*DAH7PS suggests that these key regions on the catalytic barrel involved in the communication of the allosteric information are highly similar. Indeed, these regions are generally well conserved across the Type Iβ family (Figure 3.7 C), which is consistent with the involvement of these contact regions in DAH7PS catalysis. All contact regions are located on the top side of the barrel, close to the active site. Key residues of the β2-α2 and β8-α8 loops form direct hydrogen bonding interactions with the substrates, PEP and E4P, and segments of loop β1-α1 and β7-α7 appear to secure the correct positioning of those residues that interact with substrates. Together, the DAH7PS barrel appears to be a versatile platform for additional functions, which tend to be developed conveniently by hijacking the conserved features of the active site, so that allostery can be created with ease when interdomain interactions are re-established in the presence of ligand. Additionally, some conservation of residue character for the N-terminal regulatory domains (charged, polar, and non-polar) may also contribute to the

establishment of an allosteric network between DAH7PS and regulatory domain, despite very different secondary and tertiary structures of the ACT and CM domains. The linkers between the DAH7PS and its regulatory domain share a surprisingly high degree of conservation, possibly because of the conformational restrictions for the hinging motions occurring on binding the allosteric effectors.

In addition to the transfer of allostery, enzymatic activity of the chorismate mutase domain was also transferred into the ^{Gsp}CM-^{Tma}DAH7PS chimera. Remarkably, thermostability was determined by the more thermophilic component of the chimera, even when that component, as in the case of the ACT domain of the DAH7PS from the hyperthermophilic *T. maritima* was transferred onto the much larger DAH7PS domain from the less thermophilic *G. sp* – the proverbial tail wagging the dog. The generality of this phenomenon, as a mechanism for inducing thermostability into more mesophilic enzymes remains to be elucidated.

From one enzyme and two distinct regulatory domains, we have four allosterically regulated enzymes, two occurring in Nature and two created in the laboratory, one of which is additionally a new bifunctional enzyme. The engineering of allosteric control, enzymatic activities and thermostability by domain swapping illustrates that homologous $(\beta/\alpha)_8$ structures can tolerate new modular domain combinations. Future investigations into the details of the corresponding allosteric networks experimentally and computationally are important for the design and optimisation of allosteric systems with high performance.

Chapter 4. Allostery of Type II DAH7PS

Preface

The non-covalent protein complex between DAH7PS and CM from *M. tuberculosis* was previously identified by Sasso *et al*. Complex formation was shown to significantly boost the catalytic activity of CM and allow a fine-tuned allosteric network between the two proteins.[30] Whether or not the allosteric regulation displayed by *Mtu*CM-DAH7PS represents a general allosteric strategy employed by microorganisms with a Type II DAH7PS is unknown. In the absence of *Mtu*CM, *Mtu*DAH7PS by itself presents unusual synergistic regulation by aromatic amino acids without large conformational change,[64] in contrast to the regulation via a discrete ligand-binding domain in Type Iβ DAH7PS. This chapter describes characterisation of Type II DAH7PS and CM from bacterial pathogen *H. pylori* individually, followed by investigations of the potential interactions between the two proteins. Published in Protein Data Bank in 2017, the crystal structure of *Hpy*CM (PDB 6AL9) reveals unusual features in the active site architecture and provides structural clues for interactions with *Hpy*DAH7PS. The implications of this work for providing a rational target for design of novel antibiotics against *H. pylori* are yet to be explored.

4.1. Introduction

4.1.1. Allostery and drug design

Given the extensive occurrence of allosteric regulation built on intrinsic flexibility of proteins in biological systems as described in Chapter 1, it is therefore not surprising that the field of allosteric drug discovery is being actively explored.[162-163] Compared to the classical approach targeting the functional site of a protein to execute an inhibitory effect, allosteric drugs are developed to activate or inhibit a function from a site remote to the active site. Examples of several allosteric activators are reported in an excellent review.[164]

Similar to the classical orthosteric drugs, allosteric drugs form noncovalent or covalent interactions with the target molecule.[163] Their effectiveness often relies on the conformational switch between the on- and off- states of the protein. Binding of the allosteric drug to the protein is designed to shift the free-energy landscape towards the desired energy state. The level of allosteric drugs modulates the reversible functional switch between active and inactive states, rather than eliminating a function of the target completely. Examples of allosteric drugs include valium and benzodiazepines, which target the major neurotransmitter gamma-aminobutyric acid (GABA) receptor and enhance the effect of GABA.[165-166]

Features of allosteric drugs provide them with numerous advantages over orthosteric drugs that target the active site. As allosteric sites tend to be less conserved in protein families than active sites, allosteric drugs can be highly specific with lower risk of toxicity and side effects.[163] Moreover, allosteric drugs can allow direct allosteric modulation by binding to the target protein, or indirect modulation by propagation of signals among a series of proteins through

protein-protein interactions.[144, 167] This indirect allosteric modulation can potentially expand the repertoire of allosteric drugs and allow fine-tuned allosteric control.

The discovery of allosteric drugs is challenging due to many factors. Unlike orthosteric drugs targeting known functional sites, it is difficult to predict or identify binding sites for allosteric effectors.[163] Slight alteration of allosteric interaction may cause diverse downstream effects, especially for drugs designed based on indirect allosteric modulations.[168] As with all drugs, the development of allosteric drugs also faces common problems, such as drug resistance as well as metabolic and genetic variations among patients.[163]

4.1.2. Human pathogen *Helicobacter pylori*

Helicobacter pylori (*Hpy*) is a Gram-negative bacterium that inhabits the epithelial lining of the human stomach. The bacteria rely on the contact-dependent molecular communications with the host to modulate or hijack the signal transduction pathways in the host cell.[169] It is one of the most successful human pathogens colonising around half of the population worldwide. Chronic gastritis is developed often unnoticed in all carriers, but severe gastric diseases, such as gastric cancer and peptic ulcer, occur in almost 15% of the infected population.[170-171] *H. pylori* causes a high mortality rate following these severe gastric diseases. 700,000 deaths occur every year from gastric adenocarcinoma, representing the third leading cause of cancer-related deaths worldwide.[171]

Over a decade ago, the treatment for *H. pylori* infection seemed to provide high level of success. The commonly recommended 'legacy triple therapy' includes a proton pump inhibitor (PPI), amoxicillin and clarithromycin.[172] The effectiveness of the current treatments has now declined significantly to unacceptable levels in most populations, largely due to the increase in

antimicrobial resistance to clarithromycin.[172-173] The world's leading *H. pylori* scientists have pointed out the long overdue necessity to develop alternative treatments.[172]

The long co-evolution history of *H. pylori* with humans in the past 100,000 years makes it challenging to identify alternative therapies against this bacterial infection.[169] Aforementioned knockout studies and development of enzyme inhibitors of the shikimate pathway (Chapter 1.2) emphasise the necessity of the pathway in many pathogens, providing a rational avenue for design of novel antimicrobials against *H. pylori*.

4.1.3. Interactions between DAH7PS and CM

Protein-protein interactions play important role in enzyme functions. For example, formation of protein complexes may facilitate the regulation of enzymatic activities, and substrate channelling of pathway metabolites between active sites may decrease fusion time and therefore enhance the capacity of a metabolic route.[174]

Chapter 3 detailed the interactions between Type Iβ DAH7PS barrel and its covalently tethered CM regulatory domain to support allosteric regulation of DAH7PS activity by chorismate or prephenate. In addition to this inter-domain interaction, the crystal structure of a non-covalent protein complex between Type II DAH7PS and CM from *M. tuberculosis* revealed a second type of interaction between the two proteins as briefly introduced in Chapter 1.4 (Figure 1.7).[30] Interestingly, *Mtu*CM alone was reported poorly active and unregulated. Upon complex formation, CM activity was dramatically enhanced and became regulated by Phe and Tyr synergistically, whereas the *Mtu*DAH7PS activity was unaffected.[30] The more than 100-fold increase of the CM activity suggests that this complex formation may be physiologically

essential for CM function. Recently, another Type II DAH7PS from *Corynebacterium glutamicum* was also found to form a similar non-covalent protein complex with CM.[79, 175]

4.2. *Helicobacter pylori* DAH7PS

4.2.1. Cloning, expression and purification of *Hpy*DAH7PS

The gene encoding for *Hpy*DAH7PS, from *H. pylori* strain J99 genomic DNA, was ligated into pET-32a(+) and transformed into *E. coli* BL21(DE3)-pGroESL cells by former researcher Dr Celia Webby from Parker group.[176] Attempts to replicate the previously published method for the expression and purification of *Hpy*DAH7PS were initially conducted.[177] However, the protein showed low solubility and stability under the reported experimental conditions, only 10% of enzyme activity was retained after two-days storage at 4 °C attesting to the instability of the protein (Figure 4.1 A).[176] In addition, the purification protocol involved frequent changes in the net charge of the protein by both cation exchange chromatography and anion exchange chromatography. As a result, very low yield of purified protein was obtained (1 mg per litre of cell growth). A number of methods were tested to improve the protein solubility and purification efficiency. These methods included varying growth temperatures from 37 °C to 30 °C or 23 °C, decreasing IPTG concentration from 1 mM to 0.5 mM or 0.2 mM, and varying lysis conditions and buffers with addition of additives such as DTT, various concentrations of salt, divalent metals and substrate PEP. None of these appeared to effectively improve the solubility and overall yield of the protein.

Cloning of the gene encoded for *Hpy*DAH7PS into a pDEST17 vector was performed in attempt to optimise the purification procedure by avoiding the use of ion exchange chromatography and introducing a polyhistidine affinity tag (His_6) on the protein for the ease

of detection and purification. Vector pDEST17 allows the incorporation of a TEV protease-cleavable His_6 tag on the N-terminus of *Hpy*DAH7PS via Gateway® cloning technique. The sequence-verified *Hpy*DAH7PS-pDEST17 plasmid was then transformed into *E. coli* Chaperone 3 cells for expression following standard procedures described in Chapter 7.

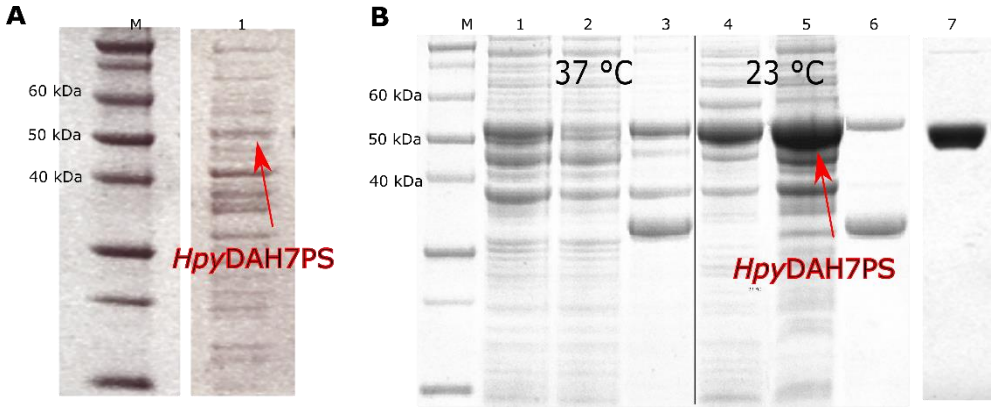


Figure 4.1. Expression and purification of *Hpy*DAH7PS. A. Soluble fraction of the unoptimised *Hpy*DAH7PS-BL21(DE3)-pGroESL construct. M, marker. 1, soluble fraction. B. Purification of the optimised His_6-*Hpy*DAH7PS-BL21(DE3)-pDEST17 construct. M, marker. 1, crude lysate of the new construct grown at 37 °C. 2, soluble fraction. 3, insoluble fraction. Lane 4-6 are repeats of 1-3, however at 23 °C growth temperature. 7, purified *Hpy*DAH7PS.

Expression trials of *Hpy*DAH7PS-pDEST17 in *E. coli* BL21*(DE3) were performed at 23 °C and 37 °C after induction with IPTG. Cell growth at 23 °C appeared to yield a high expression level of the protein with appropriate molecular weight in the soluble fraction (Figure 4.1 B). Cell lysis procedures and detailed purification methods are stated in Chapter 7. Briefly, a fast and simple purification protocol exploiting the His_6 tag on *Hpy*DAH7PS was developed. Immobilised metal affinity chromatography (IMAC) was first used to isolate His_6-*Hpy*DAH7PS from crude supernatant with mild elution conditions. Desalted protein was then treated with TEV protease to cleave the His_6 tag, which was removed by a subsequent IMAC step to produce pure *Hpy*DAH7PS. A SEC step was added when ultrapure protein was required,

such as for screening of protein crystals. A yield of 6.5 mg pure *Hpy*DAH7PS was produced per litre of cell growth, which was a significant improvement compared to the original construct. Purified protein was tested to be enzymatically active, and verified by mass spectrometry to be 51087.5 Da (theoretical value 51088.6 Da). This mass includes the complete *Hpy*DAH7PS open reading frame and three N-terminal linker residues (SGA) left over from the TEV recognition site.

4.2.2. Functional analysis

Kinetic parameters

Kinetic parameters for *Hpy*DAH7PS catalytic activities were determined following published procedures by measuring consumption of PEP at 232 nm.[177] Detailed descriptions are included in Chapter 7. The apparent K_M values for PEP and E4P were 5 ± 0.5 µM and 11 ± 1 µM respectively, and the k_{cat} value was calculated as 5.4 ± 0.1 s^{-1} (Figure 4.2).

The kinetic parameters are in line with the reported values for *Hpy*DAH7PS and *Mtu*DAH7PS (Table 4.1). Although the K_M values of *Hpy*DAH7PS for both substrates appeared to be larger than the reported values, the catalytic efficiencies remained highly consistent (1 $\mu M^{-1}s^{-1}$ for PEP and 0.5 $\mu M^{-1}s^{-1}$ for E4P). The differences between this experiment and the reported experiment include different protein constructs as stated in Chapter 4.2.1 and different experimental conditions. For example, the reported kinetic assays were performed in the presence of Mn^{2+}, however, Co^{2+} was used in this study as it is the most activating metal for the *Hpy*DAH7PS activity as reported below in Chapter 4.2.2.

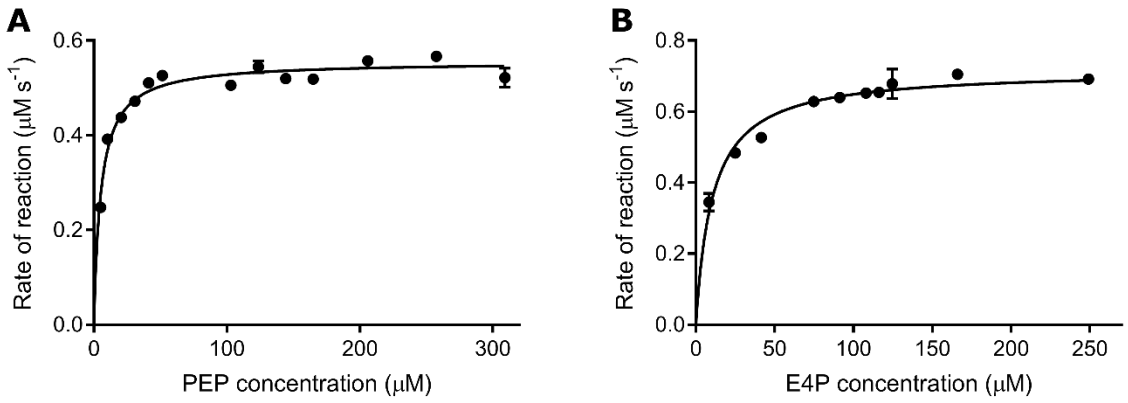


Figure 4.2. Michaelis-Menten plots for catalysis of *Hpy*DAH7PS with varying PEP (A) or E4P (B) concentrations. Concentration of PEP was fixed at 98 μM when K_M for E4P was determined. Concentration of E4P was fixed at 211 μM when K_M for PEP was determined. Error bar shows SD of triplicate measurements.

Table 4.1. Kinetic properties of *Hpy*DAH7PS and *Mtu*DAH7PS.

Organism	K_M^{PEP} (μM)	K_M^{E4P} (μM)	k_{cat} (s^{-1})	Activation by divalent metal*	Reference
H. pylori	5 ± 0.5	11 ± 1	5.4 ± 0.1	Co^{2+}, Mn^{2+}, Ni^{2+}	This study
H. pylori	3 ± 1	6 ± 1	3.0 ± 0.3	Co^{2+}, Mn^{2+}, Ca^{2+}	Webby[177]
M. tuberculosis	25 ± 1	37 ± 3	3.1 ± 0.1	Co^{2+}, Mn^{2+}, Cd^{2+}	Webby[178]

*Only the top three most activating metal ions in decreasing order of effect are shown.

Metal dependency

Several divalent metal ions were examined to identify the most activating metal for the activity of *Hpy*DAH7PS. In the presence of Co^{2+}, *Hpy*DAH7PS exhibited the highest catalytic activity, followed by Mn^{2+} and Ni^{2+} (Figure 4.3), Both Co^{2+} and Mn^{2+} provided more than 90% of *Hpy*DAH7PS activity. This is consistent with the published results for this enzyme.[177] In fact, Co^{2+} was reported to be the most activating metal ion for the activity of Type II DAH7PS in other organisms including *M. tuberculosis*, *X. campestris* and *S. caespitosus*.[178-180] However,

this does not imply that the metal dependency is type-specific. The structure of the DAH7PS catalytic core may play an important role in determining the selection of metals, as shown by two members of the Type Iβ family in Chapter 3. Large structural variations among DAH7PS members within each family should also be considered.

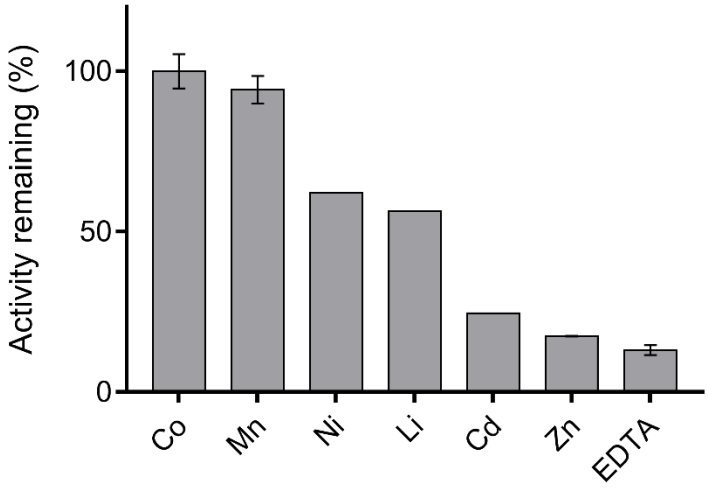


Figure 4.3. Activity of *Hpy*DAH7PS in the presence of 100 μM metal ion, or EDTA. Error bar shows SD from triplicate measurements. The mean of triplicate measurements from Co^{2+} were set as 100% activity.

pH dependency

The activities of *Hpy*DAH7PS were examined in assay buffers at a range of pH values, determined based on the theoretical pI of the protein (pI = 7.23). These values included 6.5, 7, 7.5, 8 and 8.5. The optimal pH of this protein was determined to be around 8 at experimental conditions (Figure 4.4). This pH was applied for all experiments used to characterise *Hpy*DAH7PS unless otherwise specified.

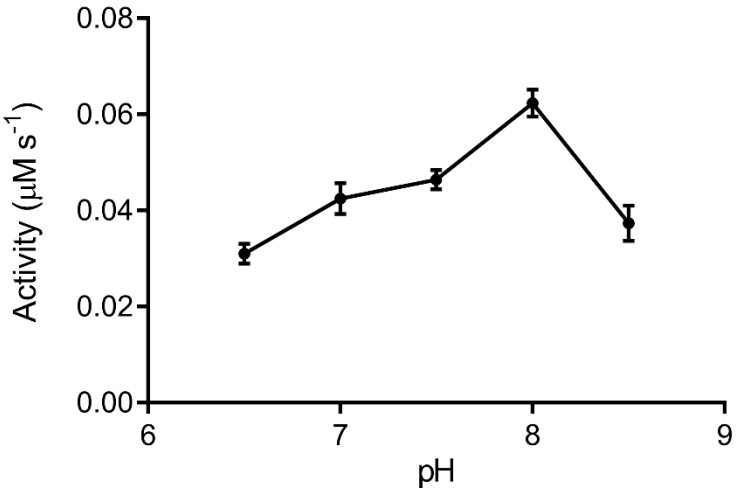


Figure 4.4. The catalytic activity of DAH7PS in response to changes in pH. Error bar shows SD from triplicate measurements.

Regulation

Kinetics - All Type II DAH7PS enzymes characterised to date are feedback regulated by aromatic amino acid(s), chorismate or prephenate.[179-184] The functionally and structurally characterised Type II DAH7PS from *M. tuberculosis* is synergistically inhibited by all three aromatic amino acids.[64] To test the sensitivity of *Hpy*DAH7PS towards aromatic amino acids, chorismate and prephenate, enzyme assays were conducted following procedures described in Chapter 7. No significant change in the *Hpy*DAH7PS activity was observed in the presence of various ligands and combinations of ligands. Different assay conditions, including different buffers (BTP, phosphate, Tris, and MES), salt concentrations (0-500 mM NaCl), and various ligand concentrations (250 μM-1 mM), were tested, no regulation of *Hpy*DAH7PS could be detected.

Table 4.2. The activity of *Hpy*DAH7PS in the presence of ligand(s). Concentration of each ligand was 250 µM. PEP concentration was 100 µM, while E4P concentration was varied (180, 50, or 15 µM). The mean activity of apo-*Hpy*DAH7PS was set to be 100%. Error represents SD of triplicate measurements.

Condition	A. Activity (%) [E4P]=180 µM	B. Activity (%) [E4P]=50 µM	C. Activity (%) [E4P]=15 µM
Apo	100 ± 0.2	100 ± 0.5	100 ± 0.5
Phe	102 ± 1	100 ± 1	86 ± 2
Tyr	110 ± 6	90 ± 6	69 ± 4
Trp	94 ± 5	92 ± 4	80 ± 1.5
Prephenate	109 ± 5	98 ± 3	95 ± 5
Phe + Tyr	101 ± 1	101 ± 1	74 ± 3
Phe + Trp	91 ± 4	86 ± 2	78 ± 0.5
Tyr + Trp	99 ± 0.5	96 ± 4	73 ± 4
Phe + Tyr + Trp	94 ± 5	97 ± 2	62 ± 1

This non-inhibitory result was unexpected, as *Hpy*DAH7PS is the only protein with homology to a DAH7PS in the *H. pylori* genome and located at a crucial node controlling the carbon flow from primary metabolism. It is also equipped with $(\beta/\alpha)_8$ barrel decorations believed to be important in ligand binding as suggested from amino acid sequence. In addition, other DAH7PS enzymes with Type II features are feedback-regulated by a single or combinations of aromatic amino acid.[64, 179-180] In *Mtu*DAH7PS, the presence of allosteric regulators alters the substrate affinity of the enzyme towards E4P.[64] Therefore, the effect of substrate concentration on regulation of *Hpy*DAH7PS was investigated. Substrate concentration was reduced from saturating level to 50 µM or 15 µM, as it is possible that the presence of ligand only alters K_M values of *Hpy*DAH7PS towards substrates while keeping V_{max} unchanged. Some inhibitory effect occurred only when E4P concentration was low (Table 4.2). The enzyme showed substantial level of inhibition when Tyr alone, or in combinations, was present in the assay with approximately 60 to 70% activity remaining. The presence of Phe and/or Trp also appeared to reduce the enzyme activity by around 10 to 20%. In contrast to *Mtu*DAH7PS, in

which combination of all three aromatic amino acid completely inhibits the DAH7PS activity,[87] the presence of Tyr, Trp and Phe only eliminated around 40% of *Hpy*DAH7PS activity.

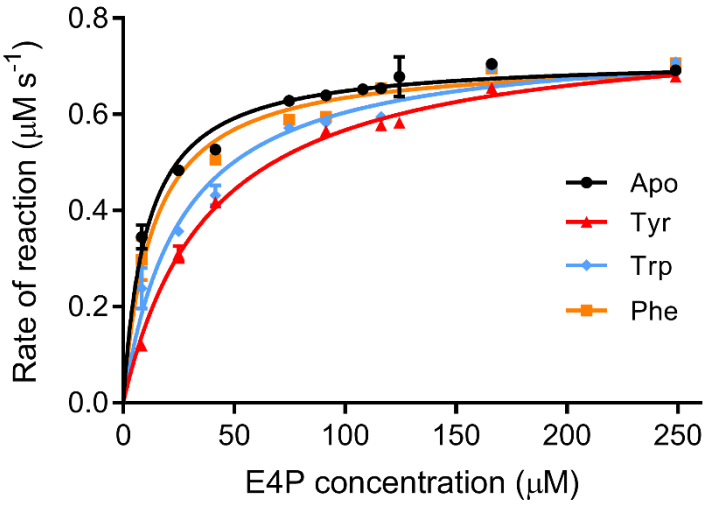


Figure 4.5. Michaelis-Menten kinetics of *Hpy*DAH7PS in the presence of individual aromatic amino acid. Concentration of PEP was fixed at 110 μM. Inhibitor concentration was 250 μM in each reaction. Error bar shows SD of triplicate measurements.

Kinetic parameters for *Hpy*DAH7PS were then determined in the presence of aromatic amino acids and compared with that of the apo-protein (Figure 4.5). The K_M value for PEP did not change remarkably with addition of individual aromatic amino acid (data not shown). However, the K_M for E4P increased around three fold in the presence of Tyr (38 ± 2 compared to 11 ± 1 μM) and around two fold in the presence of Trp (25 ± 3 μM). In the presence of Phe, the K_M only increased slightly (13 ± 2 μM), while V_{max} values were largely maintained. This result supports the non-detectable inhibition observed previously at saturating substrate concentrations. Additionally, it was reported that the presence of multiple aromatic amino acids (two or three) was able to alter the steady-state kinetic profile of *Mtu*DAH7PS towards E4P from the standard hyperbolic response to sigmoidal response demonstrating homotropic cooperativity.[64] The effect of combinations of inhibitors on kinetic parameters of *Hpy*DAH7PS

may be important for future investigations into the difference in allosteric regulation utilised by the two Type II DAH7PS enzymes.

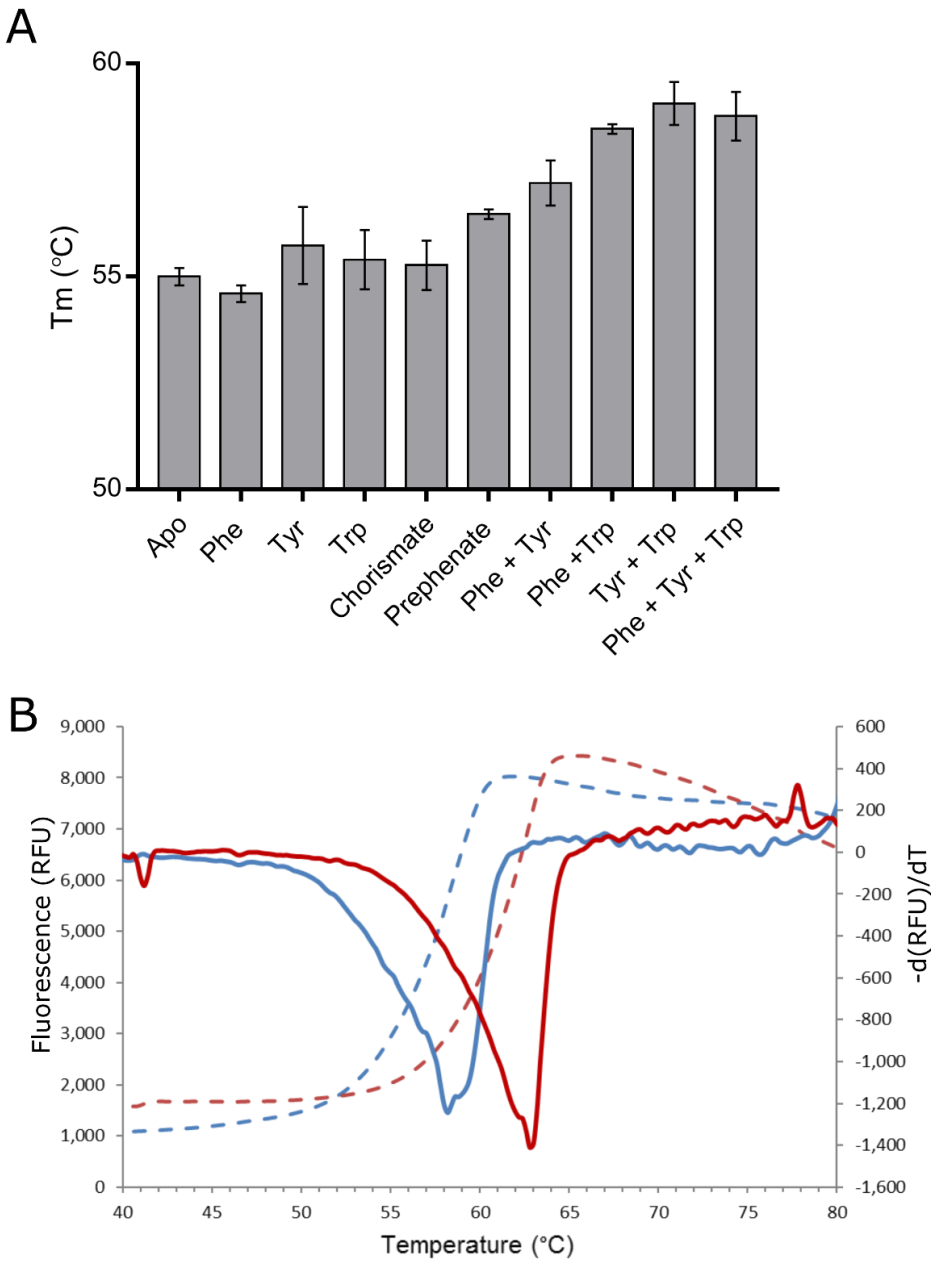


Figure 4.6. DSF analysis of *Hpy*DAH7PS. A. Thermostability of *Hpy*DAH7PS in the presence of various ligand(s). Error bar represents SD of triplicate measurements. B. Comparison between melting events of *Hpy*DAH7PS in the absence (blue) and presence (red) of three aromatic amino acids. Dashed lines show the melting curves and solid lines show the inflection points of the melting curves, which represent observed T_m values.

DSF - Given the effect of aromatic amino acids on *Hpy*DAH7PS activity, DSF was used as an initial probe to assess whether aromatic amino acids, chorismate or prephenate could bind to the protein and impact on the protein thermostability. The results showed an increase (approximately 4-6 °C) in thermostability of the protein in the presence of two, or three aromatic amino acids, compared to conditions with no ligand or single ligand (Figure 4.6). Chorismate and prephenate appeared to have no obvious effect on the thermostability of the protein. This data suggests that the protein is stabilised in the presence of aromatic amino acids, however, it does not necessarily mean that the interactions between these ligands and the protein are biologically relevant. Techniques such as isothermal titration calorimetry (ITC) could be used to verify the observations from kinetic assays and DSF, and quantify the binding affinity of the ligands towards *Hpy*DAH7PS.

4.2.3. Structural characteristics

Quaternary structure

Analytical gel filtration was used to assess the quaternary structure of *Hpy*DAH7PS (Chapter 7). The size of *Hpy*DAH7PS was determined by extrapolation from a standard curve of the log of molecular weight against elution volume/void volume (Ve/Vo) of protein standards (Figure 4.7). The calculated molecular weight of *Hpy*DAH7PS was approximately 210 kDa at 1 $mg \cdot mL^{-1}$ injection and 180 kDa at 0.2 $mg \cdot mL^{-1}$ injection.

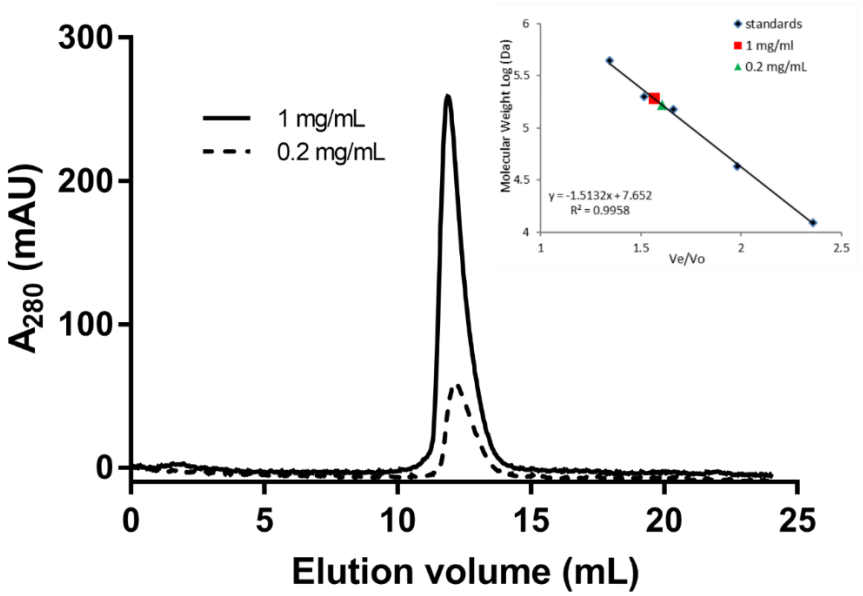


Figure 4.7. Analytical gel filtration profile of *Hpy*DAH7PS at 1 (solid line) and 0.2 (dashed line) mg·mL^{-1}.

This result is indicative of the strong presence of tetrameric species in solution under the experimental conditions. This quaternary structure is consistent with that of the Type II DAH7PS from *M. tuberculosis*,[64] however the shape of the major elution peak is not symmetrical, indicating that a mixture of species might be present with the tetrameric species being predominant. In addition, the slight right shift of the main peak at low concentration, compared to higher concentration, indicates the possible presence of smaller species. A previous publication stated that the protein was purely dimeric in solution by gel filtration,[177] however the concentration of the injected protein was not reported. This discrepancy can arise from different experimental conditions, such as concentrations, choice of protein standards, and the resolution of the analytical SEC column. To test whether *Hpy*DAH7PS exists in a true tetramer-dimer equilibrium, accurate and advanced methods including AUC and SAXS were used as described below.

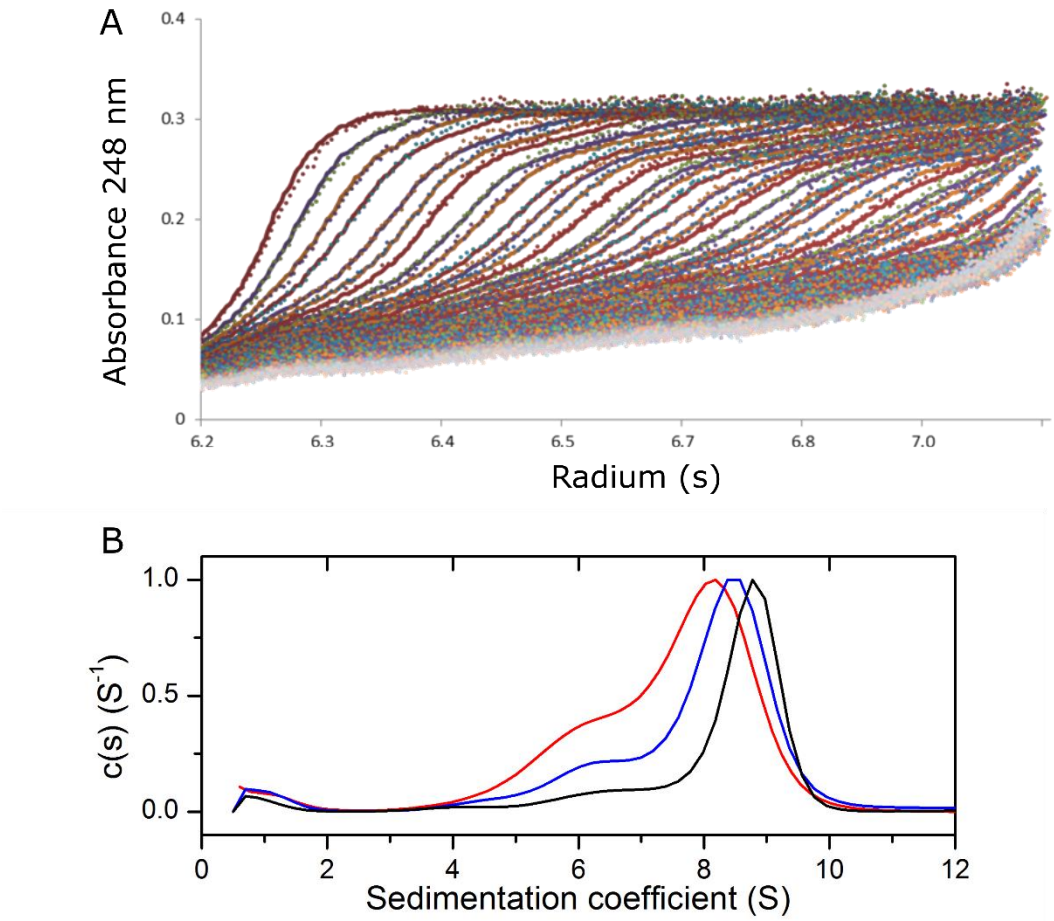


Figure 4.8. Sedimentation velocity analysis of *Hpy*DAH7PS performed at 0.9 (black line), 0.3 (blue line) and 0.1 mg·mL^{-1} (red line) at 20 °C. Data were collected at 50,000 rpm. A. The best fit (lines) of the experimental data (dots). B. Normalised distribution of the species.

Operation of the AUC instrument was performed by Dr Ali Nazmi. The absorbance versus radial position data from AUC sedimentation velocity experiments displayed multiple sedimentation boundaries (Figure 4.8 A), which suggested that two or more oligomeric populations of *Hpy*DAH7PS were present. The difference between the two major populations was confirmed by fitting the data to a continuous size-distribution model, which presented standardised sedimentation coefficients of 6 S and 8 S. The molecular mass for each population was determined to be approximately 100 kDa, and 200 kDa respectively, based on sedimentation coefficients. This was consistent with the molecular weights of the dimeric (102

kDa) and tetrameric (204 kDa) species of *Hpy*DAH7PS (Figure 4.8 B). As the protein concentration increased, the proportion of tetramer also increased relative to dimer. The shift of sedimentation coefficient towards higher values at high concentrations also supported the concentration-dependent equilibrium. Further sedimentation velocity experiments at even lower concentrations, or sedimentation equilibrium experiments, should be performed to determine if the equilibrium can be shifted towards dimer-dominant positions.

SAXS

The SAXS scattering profiles of *Hpy*DAH7PS were collected and analysed following procedures described in Chapter 7. Both reciprocal [I (s)] and real space (pair distribution function) profiles suggested an overall oblate spheroid shape of the protein. The volume and size of the protein were estimated using the structural parameters derived from SAXS (Table 4.3). The radius of gyration (R_g) determined by Guinier analysis (41.62 ± 0.03 Å) was similar to that determined by the pair distribution function (41.64 ± 0.07 Å). D_{max} was 129.77 and porod volume was estimated to be 280,256 Å^3, consistent with the theoretical volume calculated from sequence (247,256 Å^3). The molecular mass of the protein in solution was estimated to be four-times of the calculated monomeric molecular mass, confirming tetramer formation in solution.[185] OLIGOMER analysis estimated 97% of tetramer and 3% dimer in the mixture.[186] The oligomeric state of *Hpy*DAH7PS in solution determined by SAXS experiments was consistent with results from analytical gel filtration and AUC experiments. Although a concentration-dependent equilibrium between tetramer and dimer may exist, the dominant species was tetrameric at the concentrations examined.

Experimental data was then compared to the theoretical scattering data calculated from the *Hpy*DAH7PS homology model (generated by Dr Wanting Jiao using *Mtu*DAH7PS crystal structure) by CRYSOL (Figure 4.9).[187] Under the SAXS experiment conditions, the scattering

profile showed relatively poor correlation to the theoretical scattering profile calculated from the homology model (χ^2 value of 3.5). This may be an indication of a bias in the *Hpy*DAH7PS homology model towards the crystal structure of the *Mtu*DAH7PS. In addition, despite the almost 60% sequence similarity (43% sequence identity) between the two Type II DAH7PS proteins, they display important differences (Appendix D, Figure S9). The potential variations between crystal packing and solution structure may also contribute to the discrepancy between the homology model and the SAXS profile.

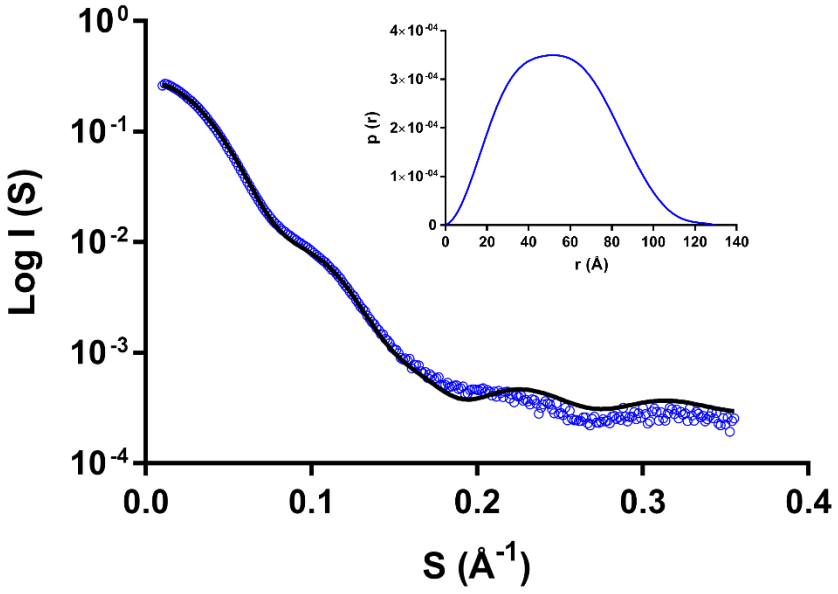


Figure 4.9. SAXS scattering profiles and P(r)-distribution of *Hpy*DAH7PS at 2 mg·mL^{-1} (blue). The SAXS data was fitted with the theoretical scattering calculated from *Hpy*DAH7PS homology model (black line).

Table 4.3. Structural parameters of *Hpy*DAH7PS obtained by SAXS.

Structural parameters	*Hpy*DAH7PS
I(0) (cm^{-1}) [from P(r)]	0.29 ± 0.0005
R_g (Å) [from P(r)]	41.64 ± 0.07
I(0) (cm^{-1}) (from Guinier)	0.29 ± 0.0003
D_{max} (Å)	126.77
R_g (Å) (from Guinier)	41.62 ± 0.03
Porod volume estimate ($Å^3$)	280,256
Dry tetrameric volume calculated from sequence ($Å^3$)	247,256
Calculated monomeric molecular mass from sequence (Da)	51,088.6
Molecular mass from Porod volume (Da) (from MoW)	206,276

Structure models

Due to the poor correlation between the homology model of *Hpy*DAH7PS with the SAXS profile, a new model was generated with the assistance from SAXS beamline scientist Dr Nathan Cowieson at the Australian Synchrotron. Rigid-body modelling was conducted using the coordinates from individual chain of the homology model to determine relative position and orientation by maximising the agreement between experimental and theoretical SAXS scatterings. For this type of modelling, each chain is considered static objects and only their relative locations are modified. Rigid-body modelling with atomic models is an efficient method to reduce fitting problem and represents a powerful and economical way to utilise information from SAXS data.[122]

The new model that correlated best with the SAXS profile (χ^2=0.73) was selected for future studies. Compared to the homology model, a significant twist at the tetramer plane was identified (Figure 4.10). The two halves of the dimer were rotated by 25° compared with the

homology model, and to accommodate this twist the two halves were also moved apart by approximately 10 Å. A previous study compared the solution scattering of *Mtu*DAH7PS with its crystal structure and suggested close correlation between the two (χ^2 = 1.24).[188] This indicates that the twist in *Hpy*DAH7PS observed from its solution state likely represents a genuine structural difference between *Hpy*DAH7PS and *Mtu*DAH7PS, although the crystal structure of *Hpy*DAH7PS is not yet determined.

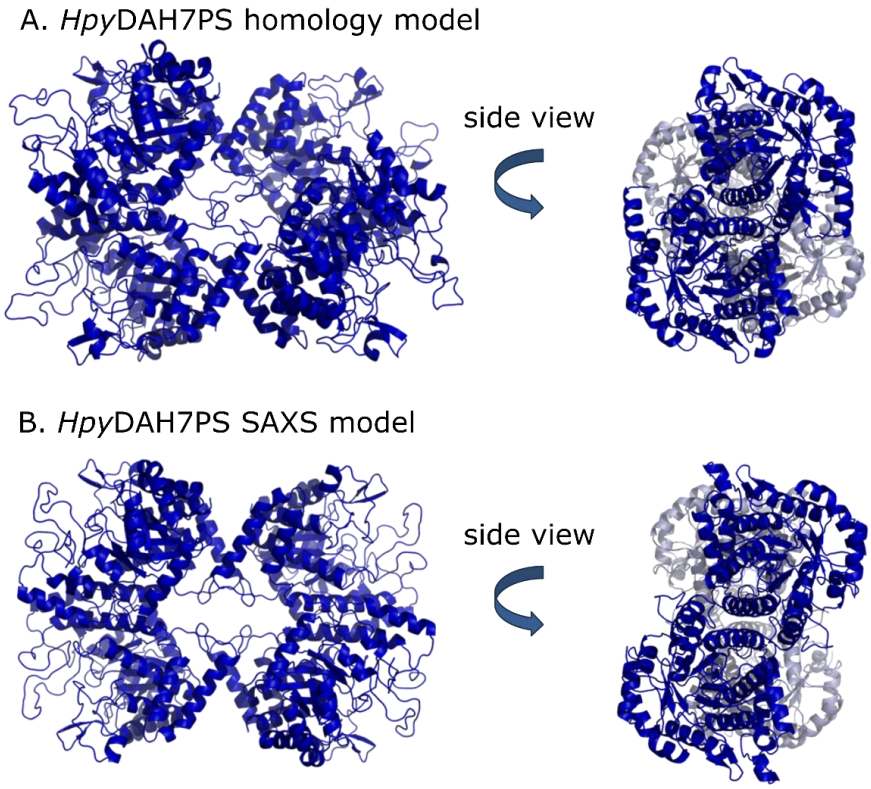


Figure 4.10. Comparison of homology model (A) and SAXS model (B) of *Hpy*DAH7PS. A twist in tetramer plane is observed in the SAXS model, indicating this protein may adopt a different shape in solution compared to the protein crystal from which it was modelled from.

Crystallisation

Extensive crystal screening of several thousand conditions was performed using highly purified *Hpy*DAH7PS in attempt to identify lead conditions for producing protein crystals. Multiple 96-

well screens from Molecular Dimensions were tested using sitting-drop vapour diffusion method, including JCSG+, PACT, Clear Strategy I and II, Proplex, MIDAS, and Morpheus (Chapter 7). Various protein conditions were also tested, including different protein concentrations (2, 4, 6, 8, and 10 $mg \cdot mL^{-1}$, severe protein precipitation occurred at concentrations higher than 10 $mg \cdot mL^{-1}$), addition of ligand and ligand combinations (PEP, E4P, Co^{2+}, Mn^{2+}, aromatic amino acids), and different incubation temperatures (5 C° and 20 C°). These attempts were unsuccessful in producing crystals with diffractions that were characteristic of protein. The equilibrium between tetramer and dimer, as well as the twisted shape of the protein in solution may contribute to the difficulty in crystallisation of this protein.

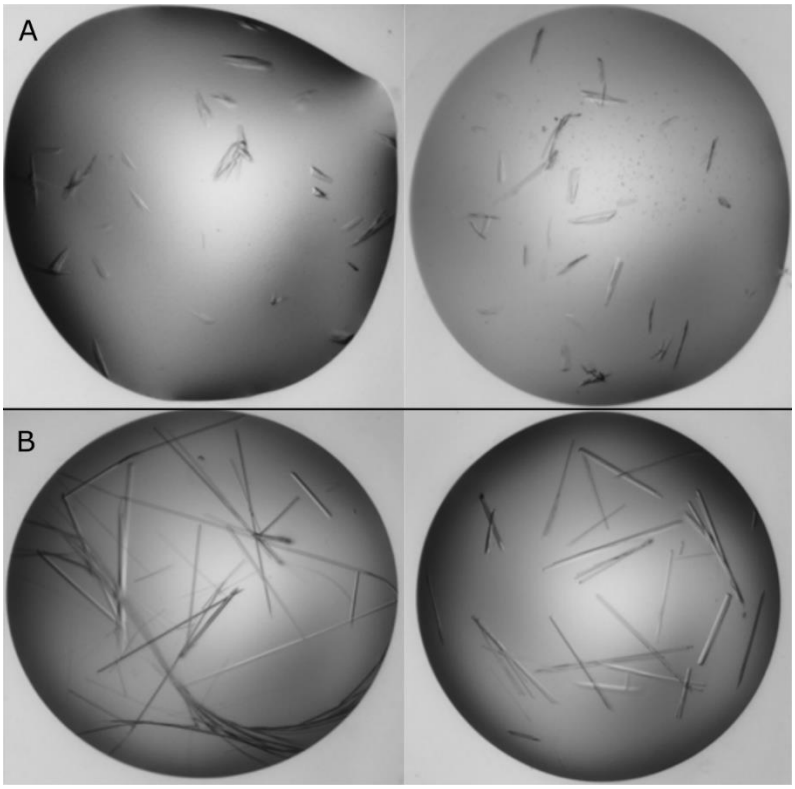


Figure 4.11. Crystals of *Hpy*DAH7PS from lead conditions.

Purified apo-*Hpy*DAH7PS (5 and 10 $mg \cdot mL^{-1}$) samples were then sent to the Collaborative Crystallisation Centre (C3) facility in Melbourne for screening of crystallisation conditions.

Two shapes of crystals were found in two conditions out of almost four hundred conditions screened (Figure 4.11). Optimisation of both conditions was performed by varying salt concentrations (0.15-0.3 M trisodium citrate or sodium sulphate) and PEG 3350 concentrations (16-26%). Only crystals with the long needle shape (Figure 4.11 B) were reproducible at low PEG concentrations (<18%). Protein samples with 5 $mg \cdot mL^{-1}$ concentration appeared to yield better crystals than 10 $mg \cdot mL^{-1}$, however, the needles were very fragile as observed when looping. These crystals were sent to both MX1 and MX2 at the Australian Synchrotron and shown to have diffractions that were characteristic of protein, although very poor. Some optimisations in crystallisation and cryoprotection conditions were performed, however, none of which improved diffraction to a workable level. Further investigation is necessary to optimise and identify conditions required to produce *Hpy*DAH7PS crystals with acceptable diffractions, as the structural information is essential in understanding the unusual structural and regulatory characteristics of this protein in molecular details.

4.3. Characterisation of *Hpy*CM

4.3.1. Expression and purification

A synthetic gene encoding for *Hpy*CM (strain 26695) was purchased from GeneArt and was codon-optimised for expression in *E. coli*. The *Hpy*CM-pDEST15-BL21*(DE3) construct was available in the Parker laboratory prior to the commencement of this project. Test expression of *Hpy*CM produced protein with the expected molecular weight by using standard protein expression techniques described in Chapter 7. The N-terminus of the *Hpy*CM construct is fused to a 26 kDa glutathione S-transferase (GST) solubility tag. A four-step purification protocol was established to isolate *Hpy*CM from impurities and the GST-tag. This purification was

achieved by firstly using a GSTrap column containing immobilised glutathione to bind to the GST-tag and isolate GST-*Hpy*CM from the crude supernatant. The resulting GST-*Hpy*CM was treated with TEV protease prior to a second GSTrap affinity chromatography step and a final SEC step (Chapter 7). The resulting *Hpy*CM was sufficiently pure (Figure 4.12). It is worth noting that the cleaved GST-tag could co-elute with *Hpy*CM at times. This residual GST-tag was removed by repeated application of the protein on the GSTrap column multiple times until the desired purity was achieved. The overall purification of *Hpy*CM yielded approximately 18 mg of protein per litre of cell culture. The size of the purified protein was determined to be 11646.2 Da by mass spectrometry, in close agreement with theoretical value of 11641.4 Da.

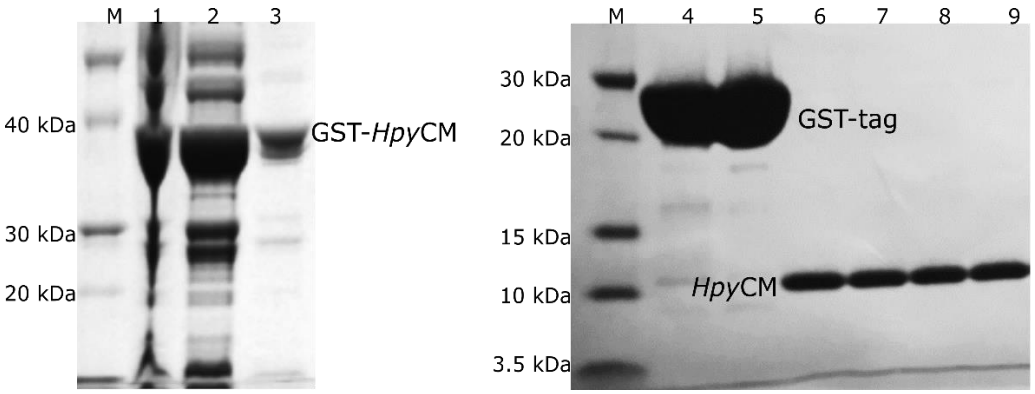


Figure 4.12. Purification of *Hpy*CM. M, marker. 1, crude lysate. 2, soluble fraction. 3, insoluble fraction. 4-5, cleaved GST tag. 6-9, purified *Hpy*CM.

4.3.2. Functional analysis

Kinetic parameters

Kinetic assays for *Hpy*CM were performed and analysed following the published procedures by measuring the consumption of chorismic acid at 274 nm at 30 °C (Chapter 7).[30] The apparent K_M value was calculated to be 94 ± 6 μM, and the k_{cat} value was 3 ± 0.1 s^{-1} (Figure 4.13). The

K_M value is in the micro-molar range, broadly in line with that of other CMs except for the *Mtu*CM, which has a K_M more than one order of magnitude higher than others (Table 4.4). The catalytic efficiencies tend to vary among different CMs, and *Mtu*CM appears to have the lowest efficiency.

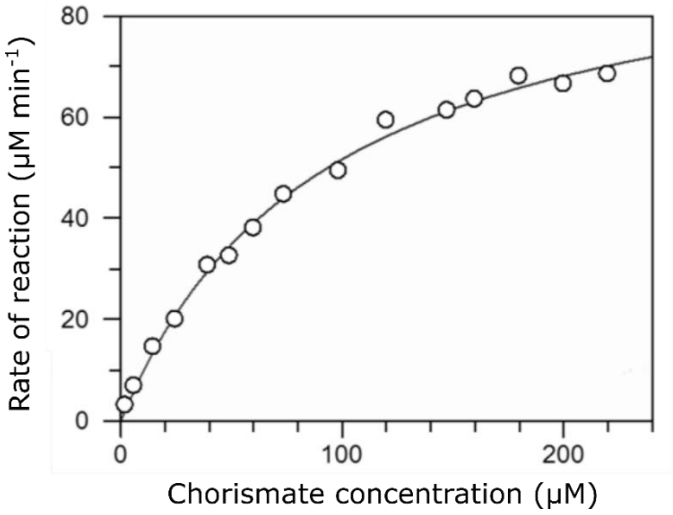


Figure 4.13. Michaelis-Menten plot for determination of the K_M value of chorismic acid for *Hpy*CM. Initial velocity values were determined as a function of chorismate concentration in BTP buffer (50 mM, pH 7.5) at 30 °C. The reaction was initiated by the addition of chorismate.

Table 4.4. Kinetic parameters of AroQ CMs from different organisms.

Organism	**K_M (μM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_M ($\mu M^{-1}\ s^{-1}$)**	**Reference**
H. pylori	94 ± 6	3 ± 0.1	3.2×10^{-2}	This study
M. tuberculosis	1140 ± 20	2 ± 0.1	1.7×10^{-3}	Sasso[30]
*E. coli**	30 ± 1	41 ± 0.4	1.4	Zhang[189]
P. aeruginosa†	98	6	6.5×10^{-2}	Calhoun[190]
S. typhimurium†	142	9	6.3×10^{-2}	Calhoun[190]
P. furiosus	25 ± 2	7 ± 0.7	2.8×10^{-1}	This study‡

*This CM is part of the *E. coli* P-protein. †Periplasmic CMs. ‡*Pfu*CM was characterised as part of control experiments in this study. Details on preparation of *Pfu*CM are included in Chapter 7.

Thermal stability

Thermal denaturation of *Hpy*CM was followed by using differential scanning fluorimetry (DSF) in 0.2 °C increments from 20 to 98 °C (Figure 4.14). The denaturation profile of the protein displays a sharp transition at approximately 51 °C followed by a subsequent denaturation event at 66 °C from the native to the complete denatured state. This biphasic melting pattern may be attributed to a two-step unfolding event, with each unfolding element occurring at a separate temperature.

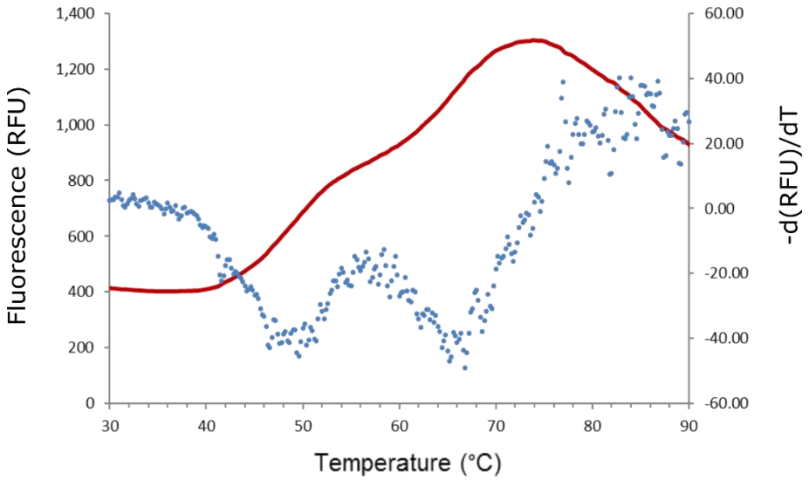


Figure 4.14. Denaturation profile of *Hpy*CM indicates a biphasic melting process. Solid line shows the melting curves and dashed line shows the inflection points of the melting curves, which represent observed T_m values.

Regulation

The AroQ CMs from different organisms appear to show either no regulation or diverse allosteric regulation by aromatic amino acids.[30-31, 115, 191-195] Inclusion of individual or combinations of Phe, Tyr, or Trp in the kinetic assays did not alter the CM activity significantly (Table 4.5). However, the presence of aromatic amino acids, especially Phe and Tyr, seemed to stabilise the protein by 2-5 °C individually, and by 7-9 °C with combinations of Phe/Tyr or

Phe/Tyr/Trp for each denaturation event measured by DSF (Figure 4.15). Trp alone did not have obvious effect on the protein denaturation. Combinations that contained Phe and Tyr exhibited similar degree of stabilisation as that for prephenate. This significant stabilisation of the protein by Phe and Tyr is intriguing, as chorismate, the end-product of the shikimate pathway, is either directed towards biosynthesis of Trp by anthranilate synthase or towards the formation of Phe and Tyr by CM. These results imply that although the aromatic amino acids do not have direct effect on the catalytic activity of *Hpy*CM, they may still interact with the protein and exhibit other roles under cellular conditions, such as, in the presence of cofactor or a protein partner.

Table 4.5. The activity of *Hpy*CM in the presence of ligand(s). Total concentration of ligand(s) in each condition was 500 μM. The mean activity of apo-*Hpy*CM was set to be 100%. Error represents SD of triplicate measurements.

Condition	Activity (%)
Apo	100 ± 1
Phe	98 ± 1
Tyr	110 ± 8
Trp	105 ± 7
Phe + Tyr	103 ± 8
Phe + Trp	100 ± 1
Tyr + Trp	107 ± 4
Phe + Tyr + Trp	109 ± 10

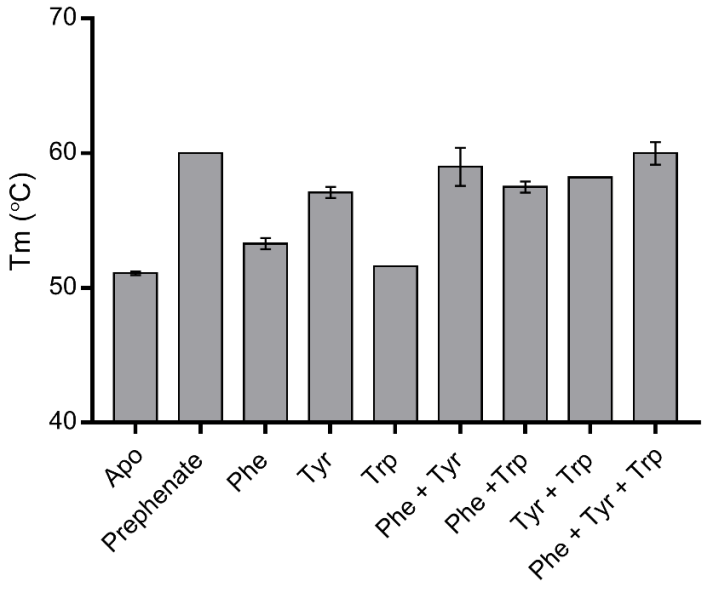


Figure 4.15. Thermostability of *Hpy*CM in the presence of aromatic amino acid(s). T_m values were recorded from the first denature event of the protein. Error bar represents SD of triplicate DSF measurements.

4.3.3. Structural characteristics

The commonly distributed AroQ-type CM typically comprises a dimer of three α-helices and is represented by the crystal structure of *Mtu*CM in complex with *Mtu*DAH7PS.[30] Thorough biochemical and structural characterisation of the *Hpy*CM were performed to aid in the investigation of the potential protein complex from *H. pylori* and the comparison with that of the *M. tuberculosis*. The solution structure of *Hpy*CM was examined mostly using techniques including analytical gel filtration, AUC and SAXS. Atomic structure was obtained and analysed using X-ray crystallography.

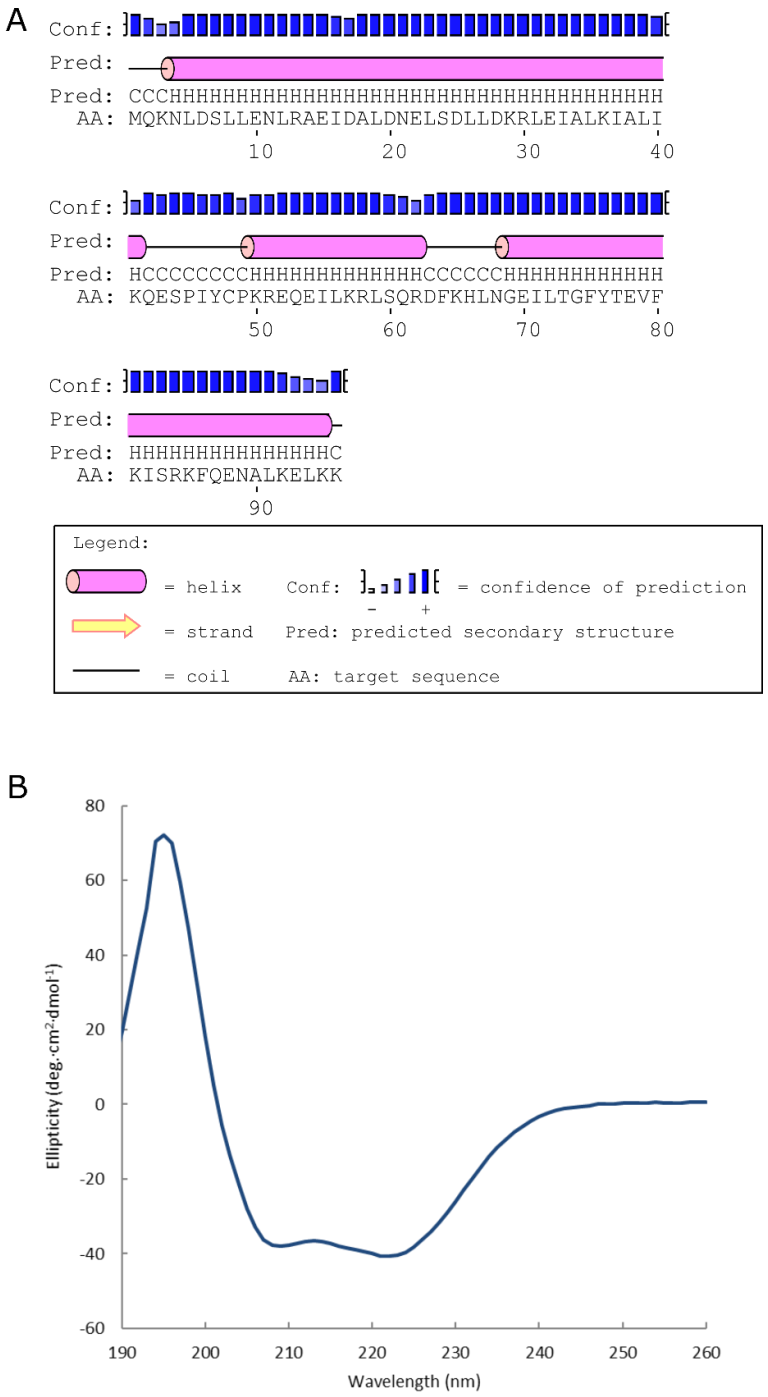


Figure 4.16. Secondary structure of *Hpy*CM. A. Secondary structure prediction from amino acid sequence using PSIPRED server. B. CD spectrum recorded from 0.1 μM protein in SEC storage buffer.

Secondary structure

The amino acid sequence of *Hpy*CM was analysed by using the PSIPRED server,[196] which predicted the secondary structure of the protein to comprise three α-helices connected by two

loop regions consistent with the typical fold of an AroQ CM (Figure 4.16 A). This structure was verified by circular dichroism (CD) spectroscopy with the observed troughs at approximately 208 nm and 222 nm, typical of α-helical proteins (Figure 4.16 B).[197] The 81% α-helix content predicted from amino acid sequence was in agreement with the 79% estimated from the CM spectrum.[198]

Solution structure

The quaternary structure of *Hpy*CM was investigated by analytical gel filtration and AUC (Figure 4.17). The molecular weight of *Hpy*CM calculated from the standard curve generated by gel filtration experiments was 24 kDa, approximately twice of the theoretical molecular mass 11.5 kDa. This result suggests that the protein is dimeric in solution under experimental conditions. This quaternary assembly was confirmed by AUC experiments. Operation of the AUC instrument was performed by Dr Ali Nazmi. Data from sedimentation velocity experiments was fitted to a continuous size-distribution model, which presented standardised sedimentation coefficients of approximately 2.2 S. The molecular mass was determined to be approximately 25 kDa based on sedimentation coefficients, in close agreement with the molecular weight of a dimeric *Hpy*CM.

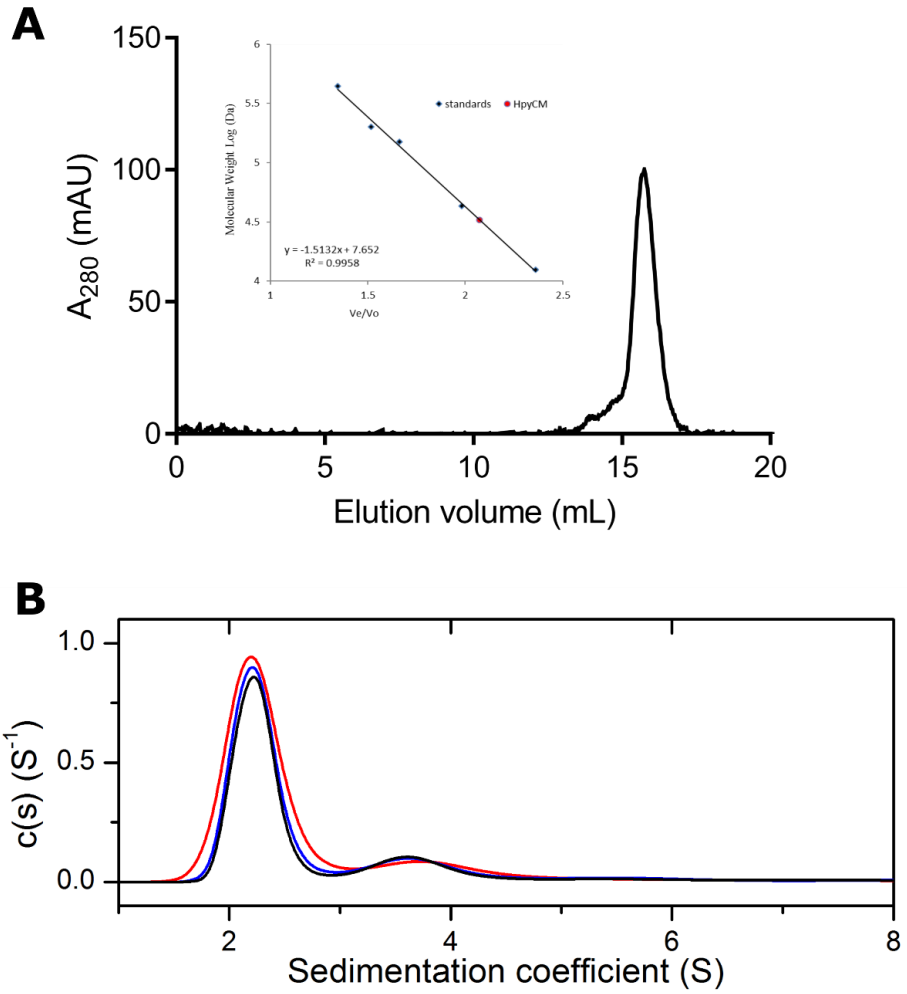


Figure 4.17. Quaternary structure of *Hpy*CM in solution is dimeric as determined by gel filtration (A) and AUC (B). A. Injected protein concentration was 1 mg·mL^{-1}. B. Normalised distribution of the species from sedimentation velocity experiments performed at 0.6 (red line), 0.4 (blue line) and 0.2 mg·mL^{-1} (black line) at 20 °C.

SAXS was used to examine the characteristics of *Hpy*CM in the absence and presence of chorismate or prephenate. The scattering data were collected with the co-flow technique at the Australian Synchrotron, and analysed following procedures described in Chapter 7. The reciprocal and real space profiles suggested an overall long rod shape of the protein (Figure 4.18 AB). Upon addition of chorismate or prephenate, the rod shape appeared to be more stretched and extended. The protein also became less flexible in the presence of ligand as

suggested by the Kratky plots (Figure 4.18 C), and smaller and more compact as suggested by the decreased R_g, $D_{max,}$ and porod volume values of the protein (Table 4.6). The SAXS MoW data also confirmed the dimeric assembly of *Hpy*CM in solution as determined by gel filtration and AUC.

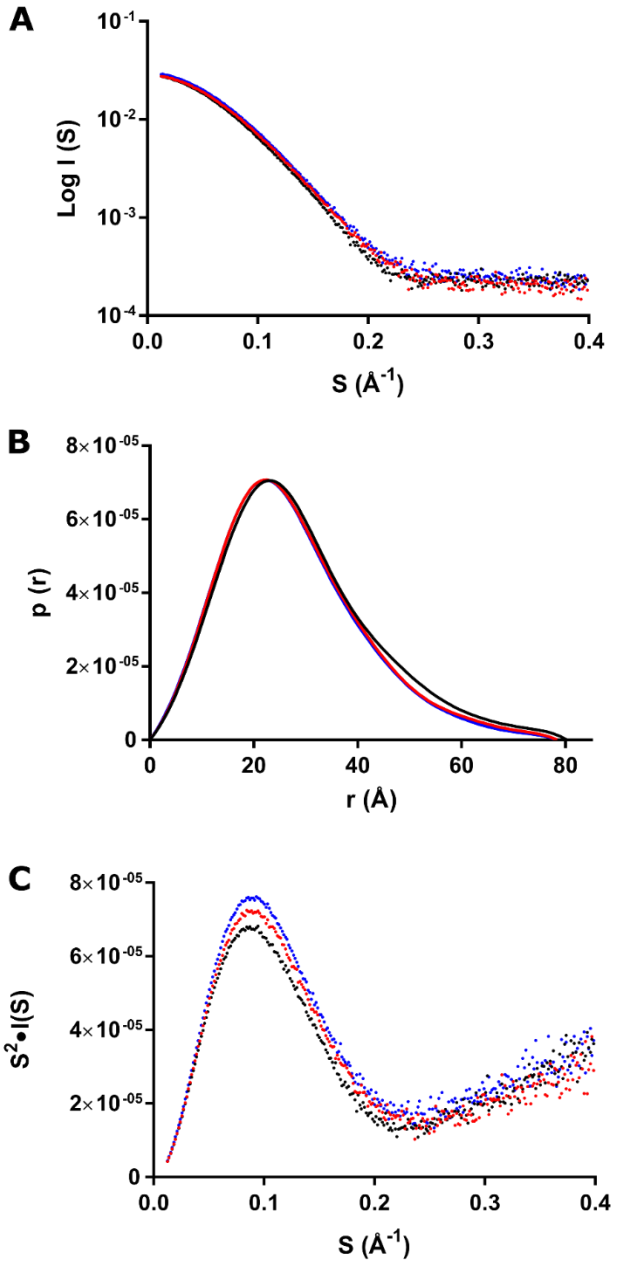


Figure 4.18. SAXS profiles of *Hpy*CM in the absence or presence of 500 µM chorismate or prephenate. A. Reciprocal scattering profiles. B. Pair distribution function. C. Kratky plots. Data from apo-*Hpy*CM

are shown in black dots or line. Conditions contained chorismate are in red. Conditions with prephenate are in blue.

Table 4.6. Structural parameters of *Hpy*CM obtained by SAXS in the absence or presence of chorismate or prephenate.

Structural parameters	Apo-*Hpy*CM	Chorismate	Prephenate
I(0) (cm^{-1}) [from P(r)]	0.28 ± 0.001	0.276 ± 0.001	0.275 ± 0.001
R_g (Å) [from P(r)]	22.83 ± 0.17	22.03 ± 0.18	21.89 ± 0.17
I(0) (cm^{-1}) (from Guinier)	0.28 ± 0.001	0.276 ± 0.001	0.275 ± 0.001
D_{max} (Å)	80.25	78.88	79.43
R_g (Å) (from Guinier)	22.81 ± 0.2	21.9 ± 0.2	21.84 ± 0.2
Porod volume estimate ($Å^3$)	46,282	41,737	41,180
Dry tetrameric volume calculated from sequence ($Å^3$)	13,969	13,969	13,969
Calculated monomeric molecular mass from sequence (Da)	11,546	11,546	11,546
Molecular mass from Porod volume (Da) (from MoW)	26,712	24,015	21,613
Number of subunits (from MoW)	2	2	2
Fit with crystal structure (χ^2) (from CRYSOL)	2.284	1.619	1.699

Crystallisation

Initial crystallisation trials for apo-*Hpy*CM at multiple concentrations were conducted using JCSG+, PACT, Clear Strategy I (CSI) and II (CSII) screens (Chapter 7). Some small and shattered crystallites were observed in several conditions, including CSI A3, C1 and G10, CSII H8, and JCSG+D6, unfortunately none of the conditions produced crystals appropriate for diffraction. Chorismate was then added to the protein at a final concentration of 500 µM, and the mixture was incubated at 4 °C for three hours prior to crystallisation screening. In addition to the aforementioned screens, Morpheus and MIDAS were also used to screen protein samples

at multiple concentrations from 5 to 22 $mg \cdot mL^{-1}$. Heavy precipitation occurred at concentrations higher than 22 $mg \cdot mL^{-1}$. Two shapes of crystals were found in MIDAS conditions D8 (0.1 M Tris, pH 8.0, 20% v/v glycerol ethoxylate) and G7 (0.2 M Ammonium acetate, 0.1 M MES, pH 6.5, 30% v/v glycerol ethoxylate), sharing the same precipitant glycerol ethoxylate (Figure 4.19 AB). However, these crystals diffracted poorly (9 Å) at the MX1 beamline largely due to their small size.

Optimisations of both MIDAS conditions were performed by varying pH, and the concentrations of protein, salt, and precipitant. Very slow crystal growth with condition G7 was observed after at least two weeks of incubation at 20 °C. No crystal was observed in condition D8. Therefore, G7 condition was treated as the key target for further optimisation. The pH value at around 6 to 6.5 was optimal. Ideal protein concentrations to produce the best crystals ranged from 8 to 11 $mg \cdot mL^{-1}$. Minimal and optimal ammonium acetate concentration was 0.2 M, as low salt content could prevent crystal growth completely. Minimal precipitant concentration was found to be 30%, with optimal concentrations ranging from 35% to 38%. Under optimal conditions, crystals formed after at least three days of incubation at 20 °C (Figure 4.19 C). Cryoprotectant was also optimised for best diffraction. Conditions including 20% PEG 400, 20% sucrose, 25% glycerol, and no cryoprotectant were tested. Both sucrose and glycerol were ideal cryoprotectant for *Hpy*CM crystals.

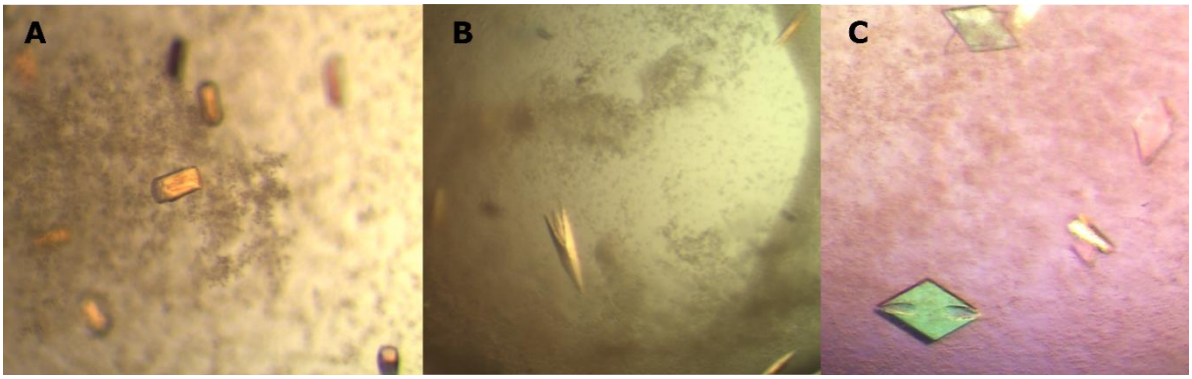


Figure 4.19. *Hpy*CM crystals. A. MIDAS condition D8. B. MIDAS condition G7. C. Optimised G7.

Table 4.7. Data collection and refinement statistics for *Hpy*CM.

	*Hpy*CM-chorismate
Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	38.87, 63.53, 166.92
α, β, γ (°)	90, 90, 90
Resolution (Å)	41.73-2.21 (2.28-2.21)*
R_{merge}	0.121 (0.372)*
$I / \sigma I$	11.6 (4.1)*
$CC_{1/2}$	0.996 (0.911)*
Completeness (%)	99.1 (99.8)*
Redundancy	8.9 (8.6)*
Refinement	
Resolution (Å)	19.43-2.30
No. reflections	9097
R_{work} / R_{free}	0.2061/0.2369
Chain length	
Protein	90 (Chain A) 91 (Chain B)
Ligand/ion	3
Water	13
Mean *B*-factors (Å)^2	
Protein	31.30
Ligand/ion	38.58 (PHB) 49.03 (PYR) 34.32 (PRE)
Water	33.15
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.301
Ramachandran	
Preferred (%)	97
Allowed (%)	3
Outliers (%)	0
PDB entry	6AL9

*Values in parentheses are for highest-resolution shell.
[Note: Wavelength of data collection, temperature and beamline are all in the section below.]

Structure determination and refinement

Three X-ray diffraction data sets were collected from different positions of the crystal at the Australian synchrotron using the MX2 beamline at 100 K following single wavelength protocol at 0.9537 Å.[199] The collected data sets were integrated and processed using XDS and AIMLESS (CCP4 program suite).[123-125] Data from two of the three data sets (collected from the ends of the crystal) showed significant twinning problem (50% by L-test, 15% by Britton). The remaining data set (collected form the centre of the crystal) appeared to be less twinned (6% by L-test, 10% by Britton) and was used for further processing and model determination. The appropriate cut-off resolution was determined via $CC_{1/2} \geq 0.5$ while ensuring the data set was complete in the highest resolution shell.[200] The data collection statistics are summarised in Table 4.7, along with key structure refinement details.

Protein *Hpy*CM was crystallised in the orthorhombic space group $C222_1$ and diffracted to 2.21 Å. Molecular replacement was performed with kind assistance from Professor Geoffrey Jameson. The monomeric structure of three search models (PDB 1YBZ, 2D8D, 3RMI with sequence identity 17.8%, 30%, 35%, respectively to *Hpy*CM) were modified by moving the loop connecting the first helix (α1) with the second and third helices (α2α3) and trimming side chains by CHAINSAW with the atoms in common with *Hpy*CM preserved.[201] The modified models were used as multicopy search models with MOLREP.[126] DM (CCP4 program suite) was used for density modifications in solvent flattening and histogram matching mode.[202] Solutions were produced with α1 and α2α3 identified consecutively by molecular replacement. TLS and restrained refinements (TLS group 1: residue 6-95/Chain A, TLS group2: residue 5-95/Chain B) were conducted with Refmac5.[203] The quality of the model was optimised by consecutive model building in Coot and refinement with Refmac5.[203-204] Water molecules were added via interpretation of the 2|Fo|-|Fc| and |Fo|-|Fc| maps, with the requirement that putative

water molecules formed at least one hydrogen bond with the protein. Electron density for residues 1-5 and 96 in Chain A, and 1-4 and 96 in Chain B was absent, which may be attributed to these residues being part of the flexible loop at both termini. The validation tools of Coot were used to assess structure quality during refinement cycles and before deposition.[204] Following refinement of the protein and water molecules, electron density of respective ligands was observed in the active sites. Ligands PRE, PHB and PYR were added and manually positioned into the electron density using Coot and then refined with Refmac5. Model elaboration and refinement continued until no further improvement of R_{work}/R_{free} could be achieved.

Crystal structure

Overall structure - The asymmetric unit contains two *Hpy*CM monomers, each composed of three α-helices connected by loops forming an intertwined tight dimer packed into the unit cell (Figure 4.20 A). Several residues in the α1α2 loop regions and the C-termini are disordered. This disordering appears to be more severe in Chain B, resulting in poorly defined α1α2 loop compared to that of Chain A. Superposition of Chain A and Chain B shows high similarity between the two chains as expected, with a root square mean deviation (RMSD) value of 0.19 Å. The only differences are situated in the active-site loop regions and the termini (Figure 4.20 B).

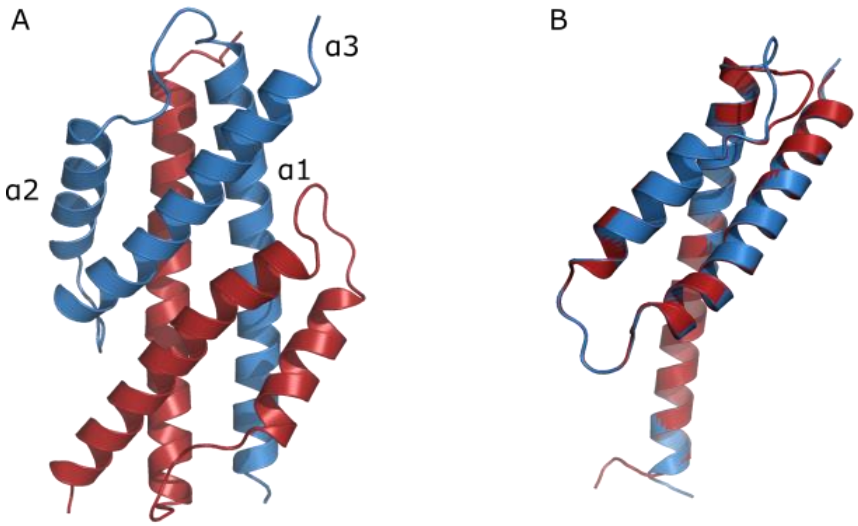


Figure 4.20. Crystal structure of *Hpy*CM. A. The protein is composed of two monomers in blue and red. B. Structural superposition of the two monomer shows differences in α1α2 loop and terminal regions.

Although the sequence identity between *Hpy*CM and *Eco*CM is only 25%, their overall structural features are comparable (Appendix D, Figure S10). The overall RMSD for the superposition of *Hpy*CM with *Eco*CM (PDB 1ECM) is 3.61 Å. The major differences between the two structures arise from the orientation of the C-terminal pair of helices (α2 and 3) compared to the N-terminal helix, the disordered active-site loop and difference in length of the C-terminal helix (Figure 4.21 A). These helices appear to be more arched in *Hpy*CM. Furthermore, superposition of available CM structures with similar fold from PDB shows substantial variations with these regions, and the differences in the shortest α-helix (α2) is especially profound as evident in Figure 4.21 B. In contrast, the positioning of α1 and 3 appear very similar among these structures.

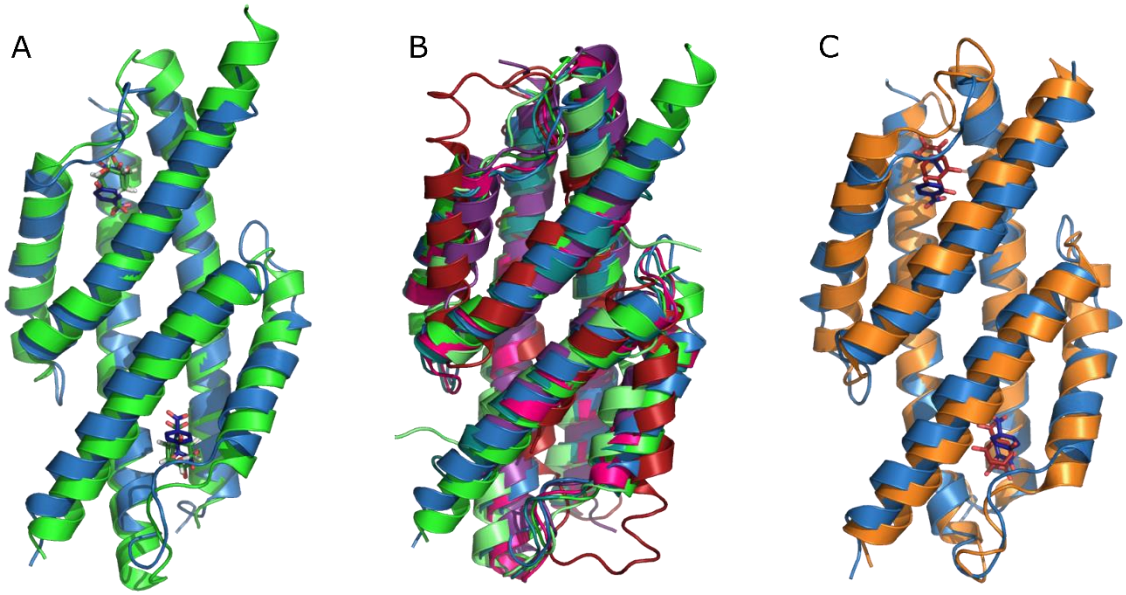


Figure 4.21. Structural comparison of *Hpy*CM with other available structures. A. Superposition of *Hpy*CM (blue) and *Eco*CM (green). B. Superposition of all CM structures with similar fold to *Hpy*CM, including PDB 1ECM, 5HUB, 2VKL, 2D8E, 2D8D and 1YBZ. C. Superposition of *Hpy*CM (blue) and *Pae*IPL (orange). Ligands at the active sites are shown in sticks.

Active site - Ligand electron density for prephenate (PRE) is identified in the active site and well-fitted in Chain A (Figure 4.22). In contrast, the density in the equivalent position of Chain B is unequivocally inconsistent with PRE, and is well-fitted by *para*-hydroxybenzoate (PHB) and pyruvate (PYR). This difference in ligand density at active site between each chain is also observed in other data sets collected under the same conditions. The difference in the active-site loops between each chain may be associated with the difference in ligand density between each active site. The active-site loop in Chain A lightly caps the active site, forming a tunnel with PRE bound in the middle and both ends opening to solvent (Figure 4.23 A). Whereas in Chain B, the active-site loop latches the entrance to the active-site pocket tightly leaving smaller solvent accessible areas and forms a relatively more 'capped' active site compared to Chain A and to the corresponding chain in *Eco*CM (Figure 4.23 B). Upon closer inspection,

the main contact associated with the tight latching of the active site is formed between the main-chain amide group of Pro45 and Ile46 in the loop region of Chain B with the carboxylate group of Glu88 from α3 helix (Figure 4.23 C). The different loop positioning in Chain A distances the sidechain of Glu88 from the loop, therefore no interaction is observed to cap the active site of Chain A exposing the active site to a larger solvent accessible area.

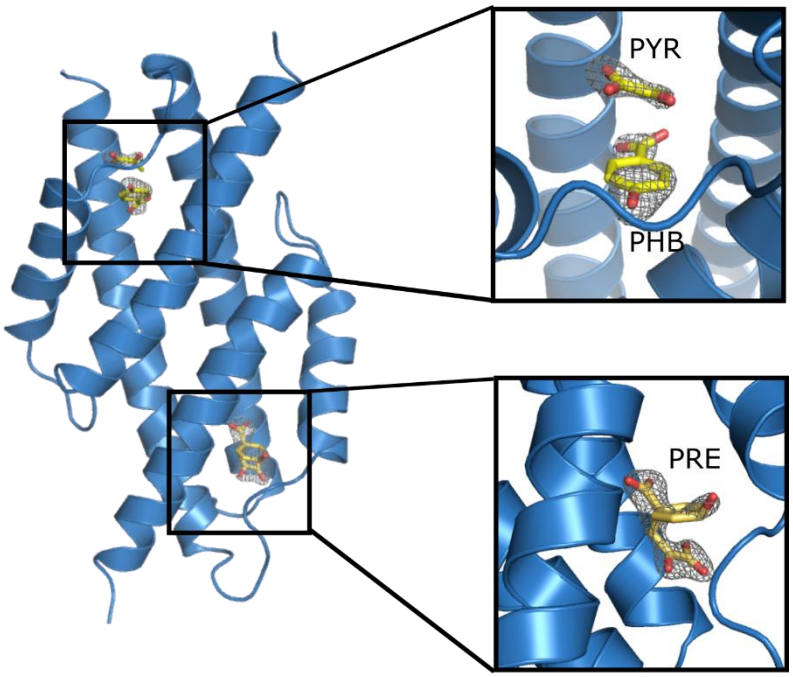


Figure 4.22. Electron density maps for the PRE, PHB and PYR bound structure of *Hpy*CM. Ligands bound at the active sites are shown as yellow sticks with their density shown by weighted |Fo|-|Fc| omit maps (grey) contoured at 3σ, generated using program FFT (CCP4 program suite) before addition of the ligand to the model.

The substrate chorismate included in the crystallisation condition is converted to product prephenate in Chain A, and to PHB and PYR in Chain B, which suggests that *Hpy*CM may moonlight as a lyase (Figure 4.24). Interestingly, the isochorismate-pyruvate lyase from *Pseudomonas aeruginosa* (*Pae*IPL), which is a structural homologue of CM, has both lyase (converts isochorismate to pyruvate and salicylate) and chorismate mutase activity (Figure 4.21

C).[205-206] This enzyme shares striking structural similarity with *Hpy*CM yielding RMSD of 2.15 Å (PDB 2H9D), albeit low sequence identity (18%). In fact, this RMSD value is lower than that with any other available CM structure (ranging from 3.19 to 9.66 Å).

Further inspection of the details of the active site shows that the PRE binding mode in Chain A of *Hpy*CM is highly similar to that of a transition state analogue (TSA), an oxabicyclic acid, in *Eco*CM (Figure 4.25 A and C).[101] In the active site of the *Eco*CM, Arg11' (from the opposing chain), Arg28, and Ser84 interact strongly with the carboxylate group at C11 of the TSA with a water molecule bridging the two carboxylates and Arg 51. Asp48 and Glu52 interact with the hydroxyl group at C4, and Lys39 interacts with the carboxylate at C11 and the ether oxygen, which also forms a hydrogen bond to Gln88. In *Hpy*CM, seven out of the eight residues important for TSA binding in *Eco*CM are highly conserved, including Arg13', Arg30, Lys41, Arg51, Glu52, Ser83 and Gln87. Six of these residues Arg13', Arg30, Arg51, Glu52, Ser83 and Gln87 form highly similar contacts with PRE in the active site of Chain A, compared to the active site of *Eco*CM. Lys41, located in the disordered active-site loop does not form direct contact with PHB in the observed crystal structure. However, it could interact with the carboxylate on C6 in alternative rotamer conformations. Also located at the active site, Tyr47 and Cys48 are not conserved. Cys48 appears to secure the correct positioning of Arg51 and Glu52. The hydroxyl group C13 of PRE also interacts with main-chain amide of Tyr47 and Cys48. Sidechain of Tyr47 is pointed away from the active site and is not involved in binding of PRE.

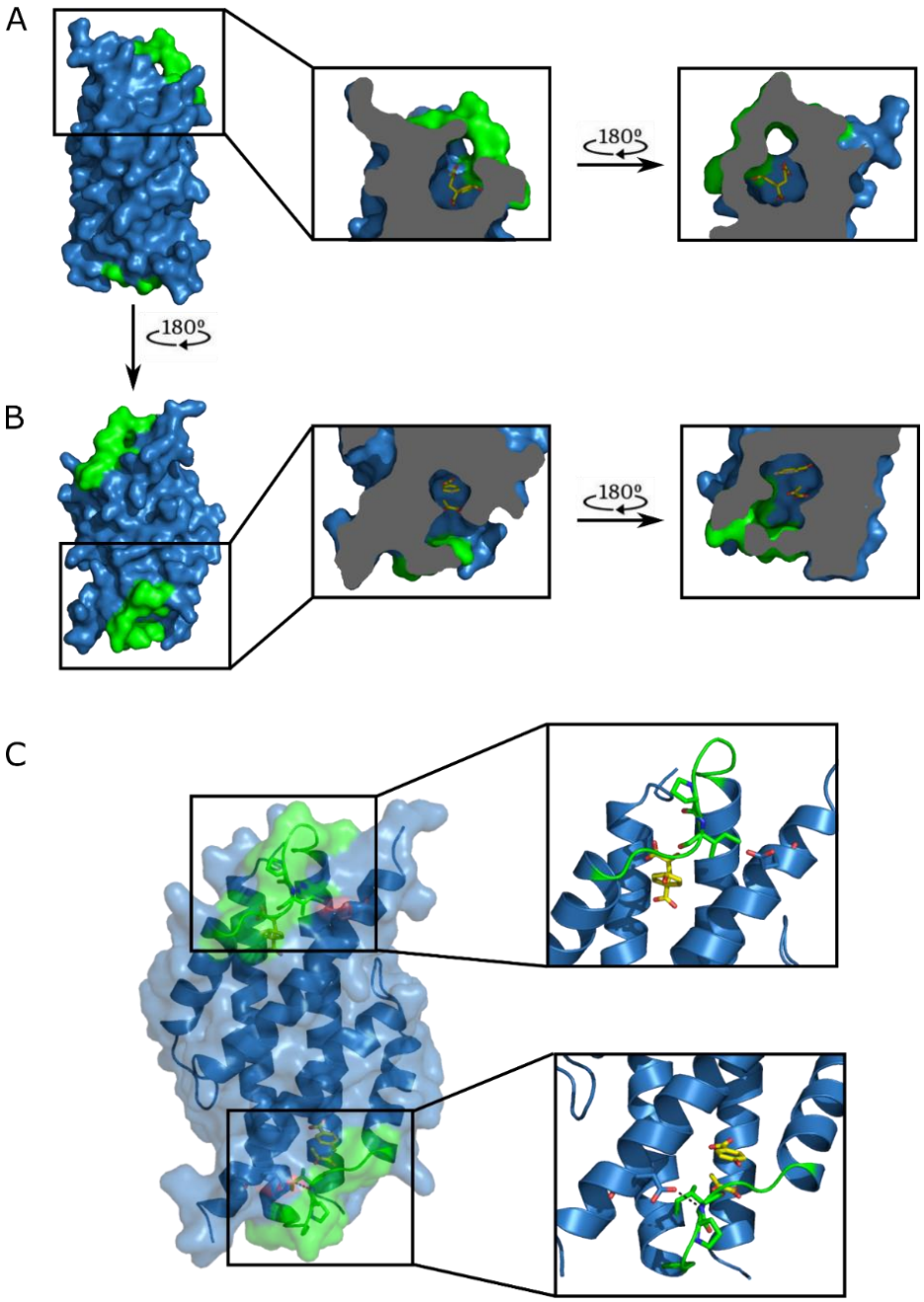


Figure 4.23. Comparison of the active-site loops in each chain of *Hpy*CM. Differences in loop arrangements cause different active-site architectures. A. Surface representation of *Hpy*CM (blue) with active-site loop highlighted in green. Cross section of solvent accessible surface showing the 'tunnel' with product PRE in the active site of Chain A. B. Active-site pocket in Chain B is less solvent accessible due to the tight latching of the active-site loop. C. Interactions between the active site loop (Pro45 and Ile46, green sticks) and α3 helix (Glu88, blue sticks) in Chain B contribute to the enclosed active site. This interaction is absent in Chain A. All ligands are shown as yellow sticks.

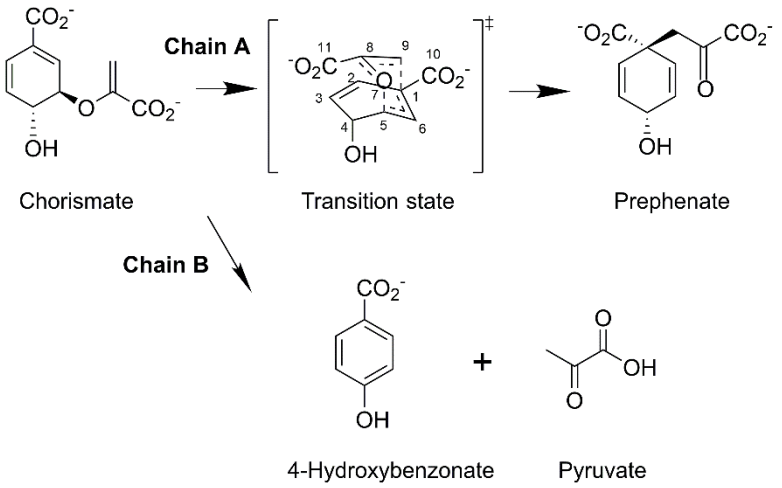


Figure 4.24. Proposed reaction scheme in each chain of the *Hpy*CM crystal. In Chain A, a typical CM catalysed reaction occurs converting chorismate to prephenate via a transition state (reaction mechanism shown in Appendix A). In Chain B, the meta pyruvyl group of chorismate appears to be removed and results in formation of 4-hydroxybenzonate and pyruvate.

The binding mode of PHB and PYR at the active site of Chain B in *Hpy*CM is highly similar to that of PRE binding to Chain A, with same residues involved (Figure 4.25 A and B). With the concomitant aromatisation of the ring, the contacts between the carboxylate group and Arg30 change slightly due to its altered orientation. The PYR binding mode is consistent with that observed in *Pae*IPL (Figure 4.25 D). PYR forms direct interactions with Arg13' and Gln87, and indirect interactions with Arg51 and PHB via the water molecule (*Hpy*CM numbering). The carboxylate functionalities of PHB in *Hpy*CM and of salicylate (SAL) in *Pae*IPL form similar contacts with the water molecule and Arg30 (Arg31 in *Pae*IPL). Additional hydrogen bonding is provided by Ser83 in *Hpy*CM. The hydroxyl functionality of PHB forms a hydrogen bond with the main-chain amide of Tyr47 and Cys48, whereas the hydroxyl group of SAL hydrogen bonds to the conserved Gln90 in *Pae*IPL. Overall, the key residues in the active sites of *Hpy*CM, *Eco*CM and *Pae*IPL are well conserved, providing the foundation for CM catalysis.

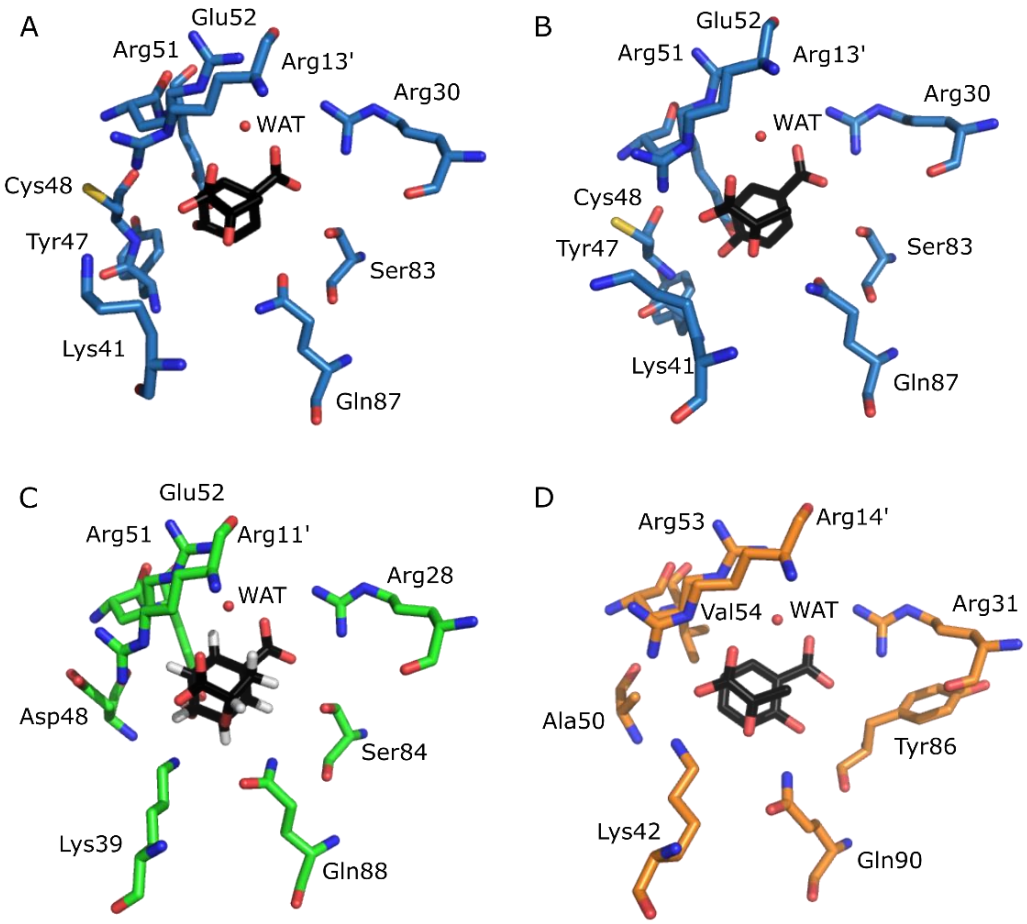


Figure 4.25. Active site constructions of *Hpy*CM, *Eco*CM and *Pae*IPL. A. Active site in Chain A of *Hpy*CM with PRE bound. B. Active site in Chain B of *Hpy*CM with PHB and PYR bound. C. Active site of *Eco*CM with TSA bound. D. Active site of *Pae*IPL with SAL and PYR bound. Key residues that form the active site are shown as sticks, active site ligand is shown as black sticks. Water molecules (WAT) are shown as red spheres.

The reduced charged interactions in the active site of *Hpy*CM compared to *Eco*CM, combined with the unusual positioning of the disordered loop may explain the active site promiscuity that probably allows the lyase activity to occur under the crystallisation conditions. This may be supported by the fact that the promiscuous active site of *Pae*IPL also has reduced charged/polar contacts compared to that of *Eco*CM. Furthermore, the conserved Lys residue in the active-site loop forms a gateway for substrate entry and product release.[205] The Lys is hypothesised to be

vital in electrostatic stabilisation of the transition state of the CM reaction, a single amino acid substitution of the Lys to a His in *Pae*IPL maintained the active site architecture but completely perished the CM activity, signifying the importance of electrostatic strength in the active site for CM catalysis.[206] In *Hpy*CM, this conserved Lys residue unexpectedly does not form obvious contact with ligand due to its unusual rotameter positioning observed in the crystal, which may alter the electrostatic field of the active site and affect CM catalysis.

Comparison with solution structure – The crystal structure of *Hpy*CM in complex with PRE, PHB and PYR was compared with the SAXS scattering profile of the protein in the absence or presence of chorismate or prephenate (Figure 4.26 A). Scattering profile calculated from the crystallographic data was best fitted with the experimental scattering of the protein in the presence of prephenate with χ^2 of 1.619 (Table 4.6). Fitting of data from the sample with chorismate was similar to that of prephenate, with a slightly higher χ^2. In contrast, scattering data from apo-protein was not well fitted with the theoretical scattering. Although the crystal structure appeared to resemble the ligand-bound SAXS data closely, some differences may reflect the effect of crystal packing on the protein structure. As evident in Figure 4.26 B, the protein from crystallographic data appeared more rigid than any sample in solution. A slightly more defined curvature in the theoretical scattering of the crystal data was also apparent (Figure 4.26 A), suggesting that crystallographic constraints may affect the observed *Hpy*CM structure, and the actual biological structure of this protein is possibly more relaxed than the crystal structure.

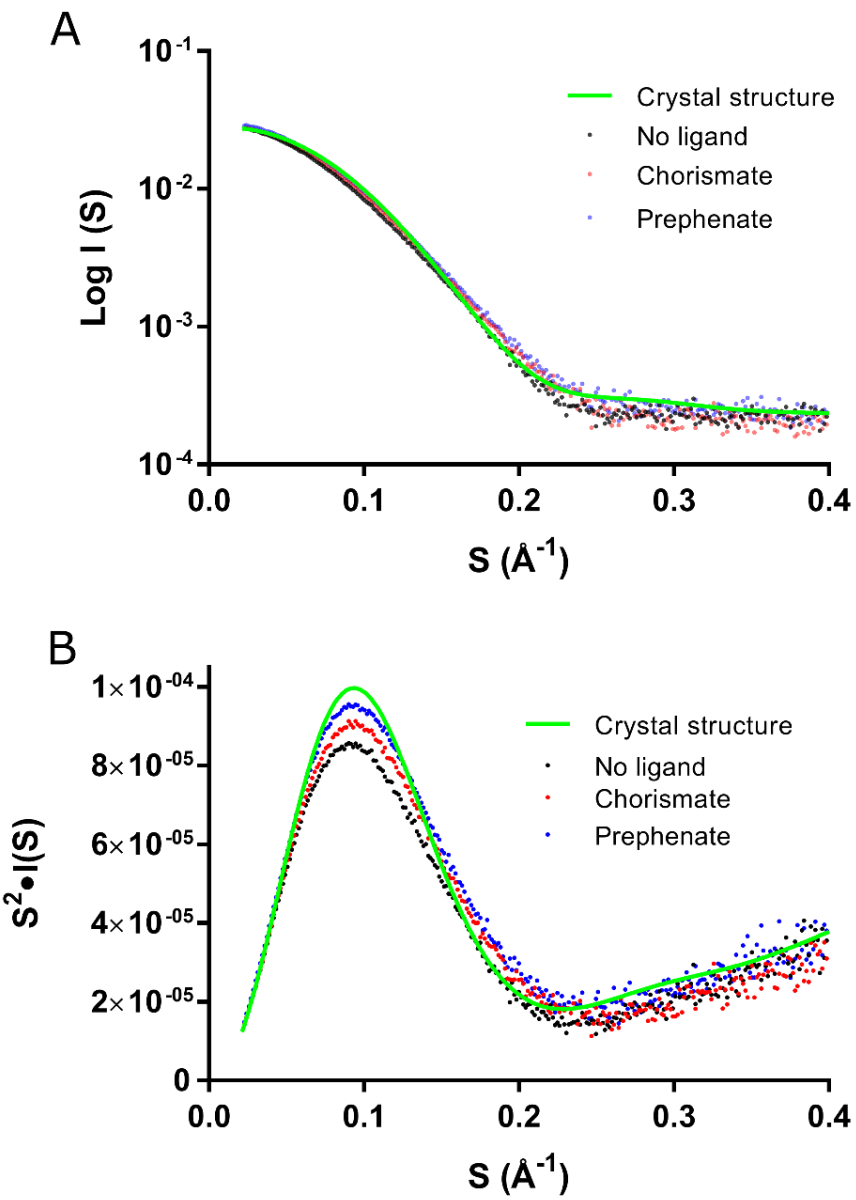


Figure 4.26. Comparison between crystal structure and solution structure of *Hpy*CM. A. SAXS scatterings of *Hpy*CM in the absence (black dots) or presence of chorismate (red dots) or prephenate (blue dots), fitted with the theoretical scattering calculated from crystallographic data (green line). B. Kratky plots show that flexibility of the protein in the crystal form closely resembles the solution structure with prephenate present.

4.4. Interaction of *Hpy*DAH7PS and *Hpy*CM

4.4.1. Protein complex detection

Several techniques were tested in attempt to capture the complex between *Hpy*DAH7PS and *Hpy*CM in solution and in crystalline form. Due to the small size of *Hpy*CM in relation to the putative protein complex and the limited resolution of some techniques, it was difficult to differentiate the complex from the *Hpy*DAH7PS tetramer. Preliminary gel filtration experiments showed a slight left shift of peak position towards larger species in *Hpy*DAH7PS sample preincubated with *Hpy*CM, compared to *Hpy*DAH7PS alone (Figure 4.27). The calculated mass for the putative complex was approximately 220 kDa, although larger than the uncomplexed *Hpy*DAH7PS (~204 kDa), it was still much smaller than the theoretical mass of the complex (~251 kDa). Therefore, a definite conclusion could not be drawn from this experiment, although the peak corresponding to the largest species might contain a mixture of complexed and uncomplexed *Hpy*DAH7PS.

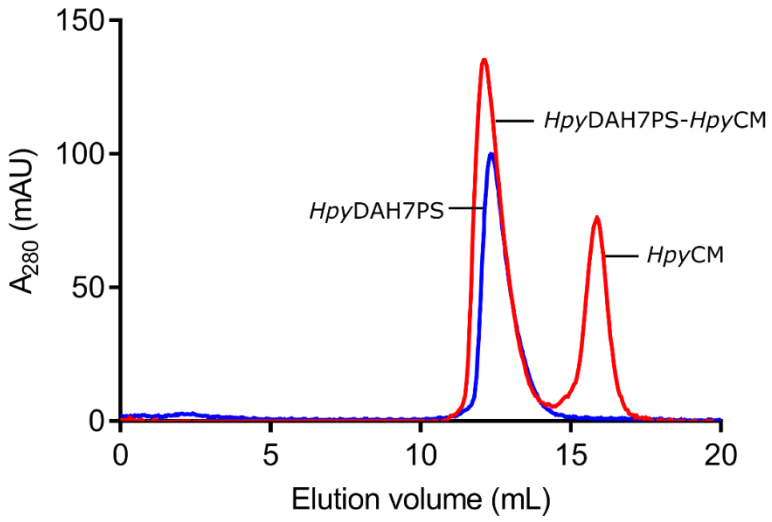


Figure 4.27. Analytical gel filtration results of *Hpy*DAH7PS with *Hpy*CM. Blue line shows sample that contained *Hpy*DAH7PS alone, red line shows a mixture of *Hpy*DAH7PS and *Hpy*CM at 1:8 molar ratio.

An IMAC pull down assay exploiting the His_6 tag on *Mtu*DAH7PS was successful in detecting the formation of complex between *Mtu*DAH7PS and *Mtu*CM.[207] Similar experiments were designed to assess potential complex formation between *Hpy*DAH7PS and *Hpy*CM. The experimental setup was intended to allow the binding of His_6-*Hpy*DAH7PS to the IMAC column and subsequently applying *Hpy*CM on the column, if complex formation occurred, *Hpy*CM would be bound to and co-eluted with *Hpy*DAH7PS. The His_6-*Hpy*DAH7PS was purified following same procedure as for *Hpy*DAH7PS, with the exception that the His_6 tag was retained. The protein was applied to an IMAC column and unbound protein was washed off before *Hpy*CM was added in the column. After half-hour incubation, unbound protein was washed off and bound protein was eluted with elution buffer.

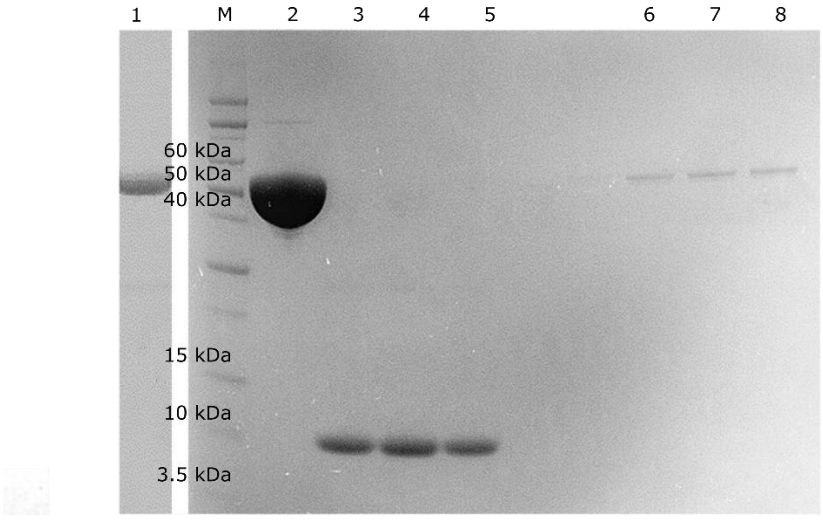


Figure 4.28. SDS-PAGE for IMAC pull down assay. M, marker. 1, His_6-*Hpy*DAH7PS loaded on the column. 2, purified His_6-*Hpy*DAH7PS. 3-5, washed fractions from IMAC. 6-8, eluted fractions from IMAC.

The SDS-PAGE (Figure 4.28) containing the unbound fractions showed a single band at approximately 10 kDa, consistent with the monomeric mass of *Hpy*CM. The eluted fractions contained a single species between 50 and 60 kDa, consistent with the monomeric mass of

His$_6$-*Hpy*DAH7PS. This result indicated that all *Hpy*CM was washed off and not complexed with *Hpy*DAH7PS. Only *Hpy*DAH7PS was bound to the column, thus no evidence of complex formation was identified from this experiment.

Table 4.8. SAXS parameters for *Hpy*DAH7PS-*Hpy*CM experiments. The sample column shows the molar ratio of each protein used (*Hpy*DAH7PS is shown as 'D', *Hpy*CM is shown as 'C').

Sample	χ	MW (Da)	R_g (Å)	Complex (fraction)	*Hpy*DAH7PS (fraction)	*Hpy*CM (fraction)
0D:10C	1.33	5715	29.49	0 ± 0	0.185 ± 0.003	0.815 ± 0.012
1D:9C	1.23	11162	32.42	0.09 ± 0.033	0.372 ± 0.04	0.538 ± 0.013
2D:8C	1.22	14543	33.08	0.119 ± 0.027	0.518 ± 0.031	0.363 ± 0.01
3D:7C	1.2	16575	33.32	0.115 ± 0.022	0.631 ± 0.026	0.254 ± 0.008
4D:6C	1.05	17391	33.46	0.195 ± 0.018	0.583 ± 0.021	0.222 ± 0.007
5D:5C	1.23	18750	33.57	0.181 ± 0.017	0.671 ± 0.02	0.148 ± 0.006
6D:4C	1.06	19072	33.64	0.239 ± 0.014	0.622 ± 0.017	0.139 ± 0.005
7D:3C	1	19569	33.68	0.257 ± 0.013	0.627 ± 0.015	0.115 ± 0.005
8D:2C	1.12	20334	33.72	0.239 ± 0.012	0.689 ± 0.014	0.072 ± 0.004
9D:1C	1.17	20799	33.76	0.244 ± 0.011	0.708 ± 0.013	0.048 ± 0.004
10D:0C	1.15	20975	33.79	0.274 ± 0.01	0.684 ± 0.012	0.043 ± 0.004

Static SAXS experiments were also designed with various molar ratios between *Hpy*DAH7PS and *Hpy*CM to determine whether complex formation could be detected. Ratios of 0:10, 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1 and 10:0 of *Hpy*DAH7PS and *Hpy*CM with total concentration of 20 µM were applied for serial sample preparation in attempt to observe a peak of complex at close to equimolar *Hpy*DAH7PS and *Hpy*CM.

Singular value decomposition (SVD) analysis was first used to analyse the whole concentration series and determine the number of independent components in the system. Three distinct

species were theoretically expected including *Hpy*DAH7PS, *Hpy*CM, and the complex. Four species were experimentally identified to be above the significance threshold (0.002) for the calculated files. Two were highly significant (1.69 and 0.02) and the other two were barely above the level for significance (0.0029 and 0.0024).

ATSAS program OLIGOMER was then used with coordinates from crystallographic data of *Hpy*CM, rigid body model of *Hpy*DAH7PS generated from previous SAXS experiments (Chapter 4.2), and homology model of the complex.[186] It is worth noting that due to signs of aggregation observed in the *Hpy*CM data, the program automatically incorporated around 18.5% *Hpy*DAH7PS to account for the aggregates (Table 4.8). Nevertheless, the decreasing trend of *Hpy*CM was still observed across the series as well as an invert trend on *Hpy*DAH7PS, consistent with the concentration series applied. A steady increase in both *Hpy*DAH7PS and the complex was observed, possibly suggesting the similarity between the two species was so high that it was difficult to differentiate them. No peak of complex was observed; hence no evidence of complex formation was identified.

The lack of evidence for complex formation is also evident by analysis of the P(r) functions from the SAXS data (Figure 4.29). Theoretical profiles of *Hpy*DAH7PS alone and the protein complex were calculated and compared with that from experimental data. The experimental profile most different from the *Hpy*DAH7PS alone was chosen for comparison. There was a subtle shift of the experimental curve to the right of the calculated curve from *Hpy*DAH7PS alone, however this change was not significant compared to the shift of the theoretical curve of the complex from that of the *Hpy*DAH7PS alone (~5 Å). Therefore, it is sensible to conclude from this experiment that the majority of the sample did not contain protein complex formed between *Hpy*DAH7PS and *Hpy*CM, and approximately 5-10% material might be in the

complex form. Such insignificant signal was not sufficient to differentiate the protein complex from *Hpy*DAH7PS alone.

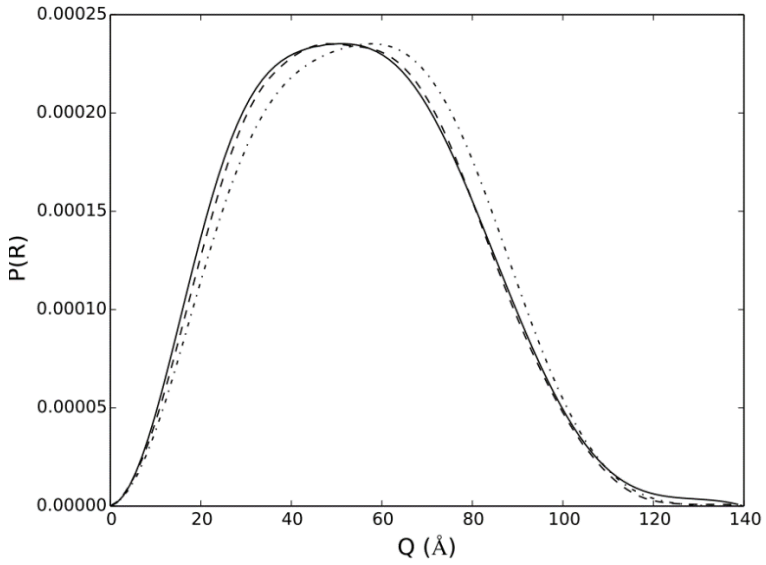


Figure 4.29. Calculated and experimental P(r) distribution of *Hpy*DAH7PS and the complex. Theoretical profile of *Hpy*DAH7PS is shown as a solid line, the complex is shown as dot-dashed line. Dashed line represents the 5D:5C experimental data.

4.4.2. Effect on *Hpy*CM

Structural investigations described above did not provide sufficient support for the formation of the protein complex between *Hpy*DAH7PS and *Hpy*CM, functional studies were performed to examine the effect of the protein mixture on each individual protein. In the *Mtu*DAH7PS-*Mtu*CM complex system, the activity of *Mtu*CM was significantly enhanced (>100-fold) in the presence of *Mtu*DAH7PS. This was believed to be a result of active site reorganisation upon complex formation in order to facilitate catalysis, while the DAH7PS activity was relatively unaffected by complex formation.[30] Similar experiments were designed to assess the effect of *Hpy*DAH7PS on *Hpy*CM and *vice versa*. The specific activity of one enzyme was determined

under a series of concentrations of the other enzyme. The presence of excess *Hpy*CM did not have significant effect on the specific activity of *Hpy*DAH7PS as determined by kinetic assays (data not shown), whereas the presence of *Hpy*DAH7PS appeared to influence the CM activity as shown in Figure 4.30.

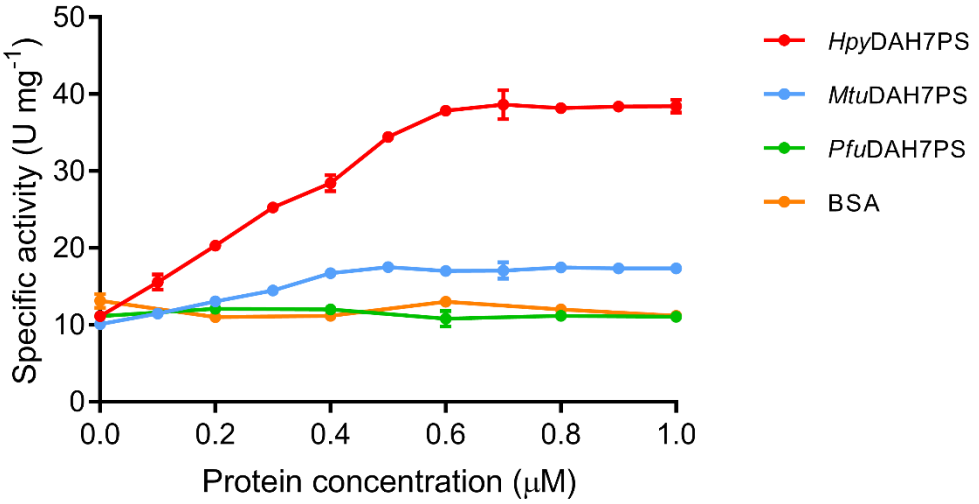


Figure 4.30. Specific activity of *Hpy*CM in the presence of *Hpy*DAH7PS, *Mtu*DAH7PS, *Pfu*DAH7PS and BSA. *Hpy*CM concentration was held constant (0.1 μM) in each reaction. Enzyme activities are specified in U (1 U = consumption of 1 μmol substrate·min^{-1}). Specific activities are given as U·mg^{-1}. Error bar represents the SD of three independent measurements. Background activity from DAH7PSs and BSA alone was subtracted from the initial rate.

In the presence of excess *Hpy*DAH7PS, the specific activity of *Hpy*CM increased by approximately 2-fold. Proteins including *Mtu*DAH7PS, *Pfu*DAH7PS (Type Iβ) and bovine serum albumin (BSA) were also tested as controls. Weak activation was observed when *Mtu*DAH7PS was present. The presence of *Pfu*DAH7PS or BSA did not cause significant change in the CM activity. Closer inspection indicated that this activity enhancement was primarily contributed by the decrease in K_M value from 94 ± 6 to 52 ± 3 μM, and slightly contributed by the increased k_{cat} from 3 ± 0.1 to 3.5 ± 0.2 s^{-1}. However, these changes were

insignificant compared to that of the *Mtu*CM, for which the K_M value dramatically decreased from 1140 ± 20 to 34 ± 3 μM, and k_{cat} increased from 2 ± 0.1 to 9 ± 1.9 s^{-1} conveying the *Mtu*CM activity to a normal range in the presence of *Mtu*DAH7PS.

Although the activation of CM activity may be linked to interaction with *Hpy*DAH7PS, it may not be physiologically relevant for two possible reasons. Firstly, compared to the weakly active *Mtu*CM, *Hpy*CM exhibits normal-range catalytic activity by itself (Table 4.4) suggesting it is functionally sufficient. Secondly, weak activation is observed in the presence of *Mtu*DAH7PS, suggesting that the Type II DAH7PS scaffolds may somehow interact with *Hpy*CM non-specifically, but not necessarily in the form of a protein complex. This intermolecular interaction may facilitate the affinity of *Hpy*CM for the substrate chorismate and influence the CM catalysis. In fact, Sasso *et al.* proposed that some homologous AroQ-type CMs would engage in activation by a Type II DAH7PS due to their consensus active-site and structural features, supported by the observation that *C. glutamicum* CM (with naturally low activity) was activated by *Mtu*DAH7PS by more than an order of magnitude.[30] Therefore, perhaps unsurprisingly, the relatively weak activation of *Hpy*CM activity was observed in the presence of *Hpy*DAH7PS or *Mtu*DAH7PS.

In addition to activation of CM activity, it has been reported that complex formation between *Mtu*CM and *Mtu*DAH7PS deliverd significant regulation to the previously unregulated *Mtu*CM by the presence of Phe and Tyr by 70% and 40%, respectively.[30] The effect of the aromatic amino acids on *Hpy*CM alone was previously examined (Chapter 4.3) and no regulation was detected, although there was an increase in thermostability in the presence of Phe and Tyr. Kinetics assays were performed with various concentrations of potential regulatory ligands to investigate if the presence of *Hpy*DAH7PS would alter the regulatory behaviour of *Hpy*CM.

The three aromatic amino acids were tested as potential feedback-regulators of *Hpy*CM individually and in combination with addition of ten-times molar excess of *Hpy*DAH7PS.

Table 4.9. Activity of *Hpy*CM in the presence of excess *Hpy*DAH7PS. Each condition contained 0.1 µM *Hpy*CM, 1 µM *Hpy*DAH7PS and 500 µM of ligand(s). The mean activity of apo-*Hpy*CM in the presence of *Hpy*DAH7PS was set to be 100%. Error represents SD of triplicate measurements.

Condition	Activity (%)
Apo	100 ± 0.8
Phe	94 ± 5
Tyr	107 ± 3
Trp	100 ± 0
Phe + Tyr	88 ± 1
Phe + Trp	99 ± 0.5
Tyr + Trp	87 ± 2
Phe + Tyr + Trp	83 ± 2

The kinetic assay showed no inhibitory effect of *Hpy*CM activity in the presence of *Hpy*DAH7PS by individual aromatic amino acid (Table 4.9). When two or more aromatic amino acids were present, weak inhibition of the CM activity within approximately 15% of its normal activity was detected. This insignificant inhibitory effect of aromatic amino acids on the *Hpy*CM activity in the absence (Chapter 4.3) or presence of *Hpy*DAH7PS indicates that physiologically *Hpy*CM is unlikely to be regulated, although it is possible that regulatory conditions, such as a protein partner and co-factor, are not yet identified.

4.5. Summary and discussion

The work described in this chapter is successful in the establishment of cloning, expression, and purification strategies for two important biosynthetic enzymes from pathogenic bacteria *H. pylori*. Both *Hpy*DAH7PS and *Hpy*CM were characterised functionally and structurally for the first time. The proposed non-covalent interactions between the two proteins were examined, however little or no complex formation was detected. While the two proteins may interact with each other, this interaction is believed to be nonspecific under the conditions tested. The physiological role (if any) of this interaction under cellular conditions is not yet known.

4.5.1. Regulation and structure of *Hpy*DAH7PS

Regulation – While *Hpy*DAH7PS and *Mtu*DAH7PS belong to the same subfamily, many differences have been identified in this study. In addition, CLAN analysis also suggests that the two enzymes belong to distinct clusters (Appendix D, Figure S11). The regulation strategy utilised by *Hpy*DAH7PS is not straight forward as described previously. A moderate level of inhibition was observed in the presence of aromatic amino acid(s), particularly Tyr, and to a reduced level for Tyr and Phe. This inhibition is only detectable at low E4P concentration due to the impact of inhibitor on K_M for E4P. In contrast, synergistic inhibition occurs in *Mtu*DAH7PS, in which Trp, Phe and Tyr are able to occupy three distinct sites formed by barrel extensions and insertions, and completely eliminate DAH7PS activity.[87] The binding mode of the inhibitors to *Hpy*DAH7PS are analysed based on the modelled coordinates in comparison with that of *Mtu*DAH7PS. Although *Hpy*DAH7PS appears to be equipped with structural barrel decorations likely associated with ligand binding (Appendix D, Figure S9), there are

fundamental differences between *Hpy*DAH7PS and *Mtu*DAH7PS, particularly at Phe/Tyr-binding sites (Figure 4.31).

Site 1 is located near the tetramer interface and residues from the inserted α2a and α2b-helices of *Mtu*DAH7PS contribute to the Trp binding site (Figure 4.31 B). This site is generally conserved in *Hpy*DAH7PS, with well-maintained pocket of hydrophobic residues for Trp binding and a few charged residues to stabilise the carboxylate functionality of Trp. As this binding site is largely contributed by the α2b-β3 loop, it is likely that Trp can be accommodated at the same site in *Hpy*DAH7PS. Indeed, a docking experiment conducted by Dr Wanting Jiao suggests that it is possible for Trp to bind in the *Hpy*DAH7PS binding pocket with additional residues such as Lys106, Arg53, and Glu229 (*Hpy*DAH7PS numbering) located nearby to potentially accommodate the carboxylate group of Trp, and the amino group of Trp can form hydrogen bonds with residues from the nearby loops in several poses. Although this binding pocket appears to be large enough to accommodate the side-chain of Trp and have the appropriate residues present, it is important to note that this site is located on the edge of the tetramer interface. In contrast to the tetrameric *Mtu*DAH7PS, *Hpy*DAH7PS is in dimer-tetramer equilibrium with the twist in the tetramer plane as described in Chapter 4.2. This difference possibly affects the Trp binding affinity at this site of *Hpy*DAH7PS, which may be associated with the relative weak inhibition of *Hpy*DAH7PS by Trp.

Phe or Tyr binding at site 2 in *Mtu*DAH7PS is capped by two β0-sheets each contributed by the tight dimer. However, sequence comparison reveals that this N-terminal β0-strand is missing in *Hpy*DAH7PS (Appendix D, Figure S9). The equivalents of some residues involved in Phe binding of *Mtu*DAH7PS can be identified in *Hpy*DAH7PS, including Pro56, Phe 91, Tyr173 and Ala178 (*Mtu*DAH7PS numbering, Figure 4.31 C). However, perhaps more importantly, Ala174 is replaced by a His residue in *Hpy*DAH7PS, Val51 is replaced by Glu,

and Val55 is replaced by Tyr. The increase in charged/polar features of this site may destabilise the binding of Phe, but promote interaction with the hydroxyl group of Tyr. Therefore, binding of Tyr at site 2 in *Hpy*DAH7PS is highly likely, and Phe may or may not interact with this site. This is supported by the relatively strong inhibition of *Hpy*DAH7PS activity by Tyr, and weak inhibition by Phe.

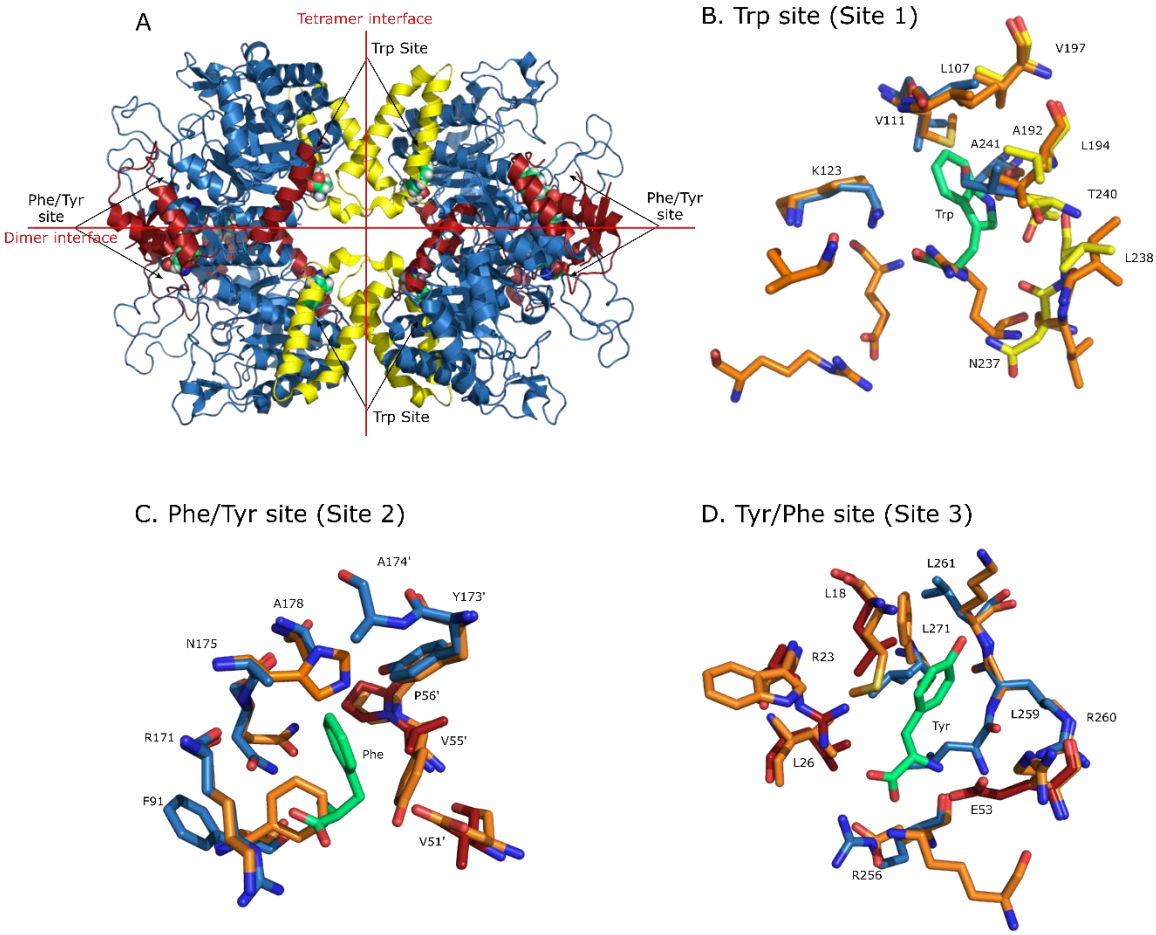


Figure 4.31. Ligand binding sites of *Mtu*DAH7PS in comparison with *Hpy*DAH7PS. A. the homotetramer of *Mtu*DAH7PS in complex with Phe, Tyr and Trp (PDB 5CKV). Vertical tetramer interface and horizontal dimer interface are indicated by red lines. The N-terminal barrel extensions (β0 and α0a-α0c helices) are shown in red, and the additional α2a, α2b helices are shown in yellow. Ligands are shown as green spheres. B. Trp binding sites close to tetramer interface with key residues shown as sticks. C. Primary Phe binding sites in the tight dimer interface. D. Tyr/Phe binding site at N-terminus.

*Mtu*DAH7PS residues that are on the N-terminal extension are shown as red sticks, the residues of the α2a-α2b helices are shown as yellow sticks, the rest are shown as blue sticks. All equivalent residues of *Hpy*DAH7PS are shown as orange sticks in B, C and D.

Site 3 is Phe or Tyr binding site nestled between the α0a and α0b-helices of the N-terminal extension and the α3-helix of the core barrel of *Mtu*DAH7PS. Residues in these regions are not conserved in the *Hpy*DAH7PS sequence (Figure 4.31 D). Only a couple of hydrophobic features can be identified at this site, and the hydroxyl, carboxylate and amino groups of Tyr (or the functional groups of Phe) cannot be sufficiently stabilised by nearby residues. Therefore, ligand binding in this site of *Hpy*DAH7PS is not favourable compared to site 1 and 2. Based on the analysis above combined with results from kinetics, inhibition assays and DSF, it is logical to conclude that feedback regulation of *Hpy*DAH7PS is substantially different from that of *Mtu*DAH7PS. A more moderate inhibition level is observed in *Hpy*DAH7PS, for instance, no complete elimination of activity is detected even when all three aromatic amino acids are present. Tyr and Phe may bind to the equivalent location of site 2, while Trp binding at site 1 is ambiguous. Crystallographic data, ITC or NMR experiments may be required to assess the ligand binding mode in detail.

Structure - Another important feature of *Hpy*DAH7PS is the quaternary structure equilibrium of dimer-tetramer as evident from AUC and SAXS results. Although minor dimeric species of *Mtu*DAH7PS occur at very low concentrations,[207] the equilibrium shifts almost entirely towards the tetramer at protein concentrations greater than 0.2 $mg \cdot mL^{-1}$, whereas dimer species of *Hpy*DAH7PS can still be distinguished at low protein concentration. The internal extensions (α2a and α2b-helices) form major contacts at the tetramer interface of *Mtu*DAH7PS. The *Hpy*DAH7PS sequence in this region appears similar to that of *Mtu*DAH7PS; however, there are still subtle changes that may increase the flexibility of this region and result in a less stable

tetramer interface. Particularly, the twist of tight dimers across the tetramer plane identified from SAXS experiments changes the relative position of each tight dimer compared to the homology model. This twist is likely associated with the lack of (β0) and/or poorly conserved (α0a and α0b) N-terminal extensions in *Hpy*DAH7PS compared to *Mtu*DAH7PS, as these regions are important in the relative positioning of each unit in the tight dimer and are likely to be crucial in determining the interaction of the tetramer interface and the overall tetramer conformation. If substantial enough, this change could be responsible for the presence of the dimer-tetramer equilibrium and even contribute to the regulatory behaviour of the enzyme. The detailed role of the N-terminal modifications of *Hpy*DAH7PS can be studied by engineering the missing β0 segment and investigating if the engineered changes at the N-terminus would alter the equilibrium and quaternary structure of the protein.

Although there are substantial differences at the dimer and tetramer interfaces of *Hpy*DAH7PS and *Mtu*DAH7PS, the lack of crystallographic data for *Hpy*DAH7PS makes it difficult to determine whether the 'twist' observed in the solution structure exists in crystalline form. Several studies have suggested that crystal-packing effects generally do not change protein folding, however, they rather affect the arrangement of flexible regions such as linkers and termini.[122] Particularly for allosteric proteins, the various sampled conformations associated with allostery are possibly affected by lattice forces. Future work on optimising the atomic model of *Hpy*DAH7PS with combination of crystallography, computational searches and SAXS is essential to fully characterise the biological conformations of this protein.

4.5.2. Structure of *Hpy*CM and complex formation

HpyCM – The crystal structure of *Hpy*CM (PDB 6AL9) reveals important features of this enzyme. While it adopts a typical AroQ-type CM fold with two tight dimers intertwined, each composed of three α-helices, it appears to moonlight as a lyase at least under crystallisation constraints. The active site promiscuity discovered in *Hpy*CM catalysis is not unique in chorismate utilising enzymes.[208-210] Although *Hpy*CM shares substantial similarity with *Pae*IPL in both overall structures and detailed active site architectures, and the PredictProtein server also reports possible lyase activity of *Hpy*CM based on gene ontology,[211] it is still difficult to conclude that *Hpy*CM moonlights as a chorismate lyase (CL) and catalyses the conversion of chorismate to PHB and PYR, as this reaction can also occur nonenzymatically.[212] It is possible that somehow in the crystallographic conditions, the CM active site is not sufficiently arranged for the conversion of chorismate from pseudoequatorial to pseudodiaxial conformation (the rate-limiting step, Appendix A), and the retention of chorismate at the active site facilitates the breakdown of chorismate into PHB and PYR. Upon completeness of the lyase reaction, PYR release is generally rapid and efficient as reported for the *Eco*CL (an enzyme catalysing the conversion of chorismate to PYR and PHB).[212] Although it is unknown whether this rapid release of PYR is maintained in *Hpy*CM, the fact that both PYR and PHB densities are present in the crystallography data is likely contributed by the covered active site in Chain B. Therefore, it is reasonable to speculate that under combined effects of crystallisation constraints and the unique electrostatic field of the active site identified in *Hpy*CM, the disordering of the active-site loop results in a more capped active site, which facilitates the retention of chorismate and allow its breakdown into PHB and PYR enzymatically or nonenzymatically. Future investigations into this lyase activity may be important in fully determining the physiological role of *Hpy*CM.

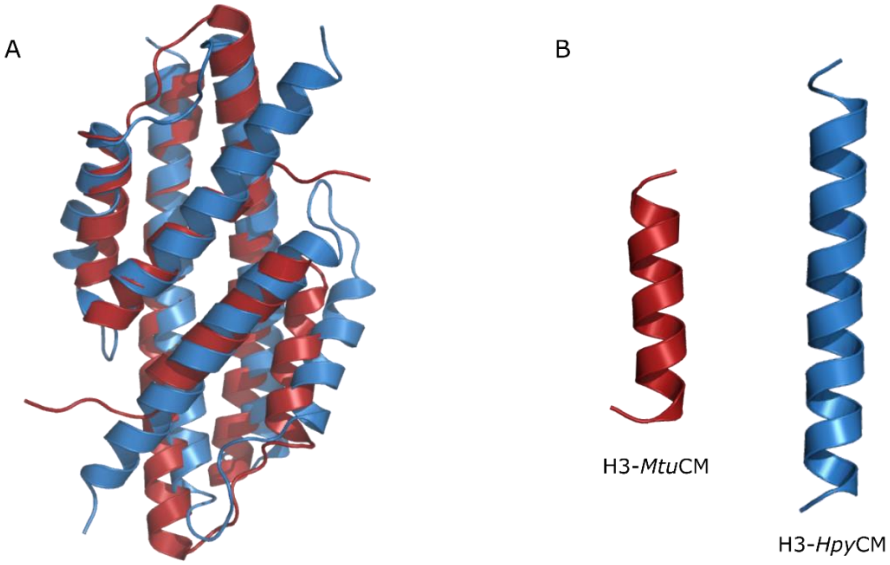


Figure 4.32. Structure comparison between *Hpy*CM and *Mtu*CM. A. Superposition of *Hpy*CM (blue, PDB 6AL9) with *Mtu*CM (red, PDB 2VKL). B. Comparison of the C-terminal helices from the two proteins.

Despite similar folding architectures, substantial differences can be identified between *Hpy*CM and *Mtu*CM. Firstly, the *M. tuberculosis* genome encodes two proteins with homology to CMs (intracellular and secreted), whereas there is only one CM in the *H. pylori* genome. This supports with the normal-range activity observed for the *Hpy*CM in contrast to the unusually weak activity of *Mtu*CM, in which a highly conserved Lys is replaced with an Arg.[30] Secondly, *Mtu*CM is strongly activated and becomes regulated when it forms a complex with *Mtu*DAH7PS; although very weak activation of *Hpy*CM activity by *Hpy*DAH7PS is detected, it does not appear to be biologically relevant or significant. Additionally, *Hpy*CM activity appears largely unregulated under all experimented conditions. Finally, unusual structural features are present in the *Mtu*CM compared to the *Hpy*CM. For instance, the C-terminal helix (α3) in *Hpy*CM is significantly longer (around two-fold) than that of *Mtu*CM (Appendix D, Figure S10, Figure 4.32). The missing segment carries a catalytically important and conserved Gln residue in other CMs, such as *Hpy*CM and *Eco*CM. The special arrangement of the short

α3 helix of *Mtu*CM plays an important role in complex formation and the mechanism of its activation by *Mtu*DAH7PS.[30]

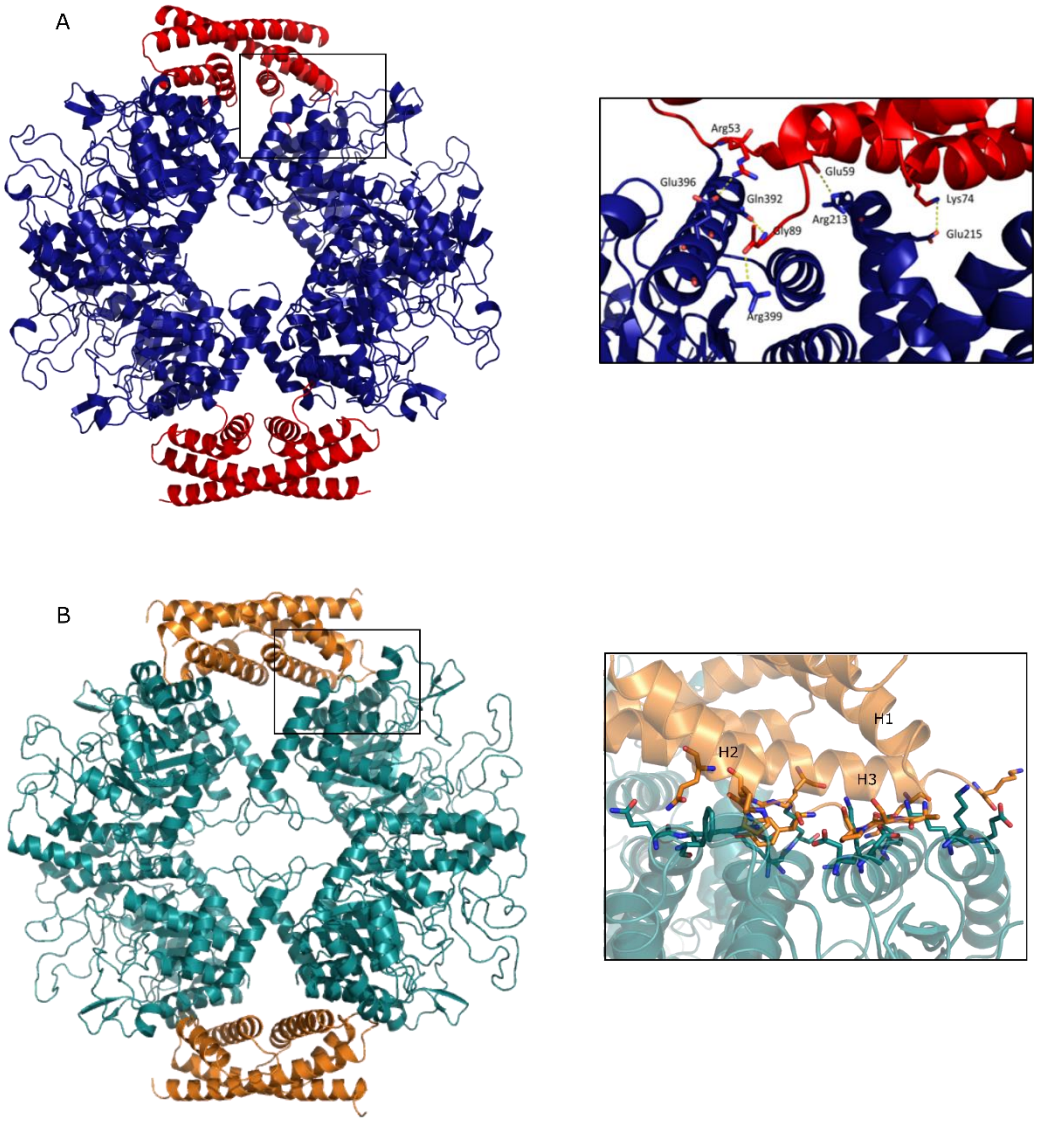


Figure 4.33. Interactions between CM and DAH7PS in *M. tuberculosis* and *H. pylori*. A. Crystal structure of *Mtu*CM (red)-*Mtu*DAH7PS (blue) complex, with key interface residues shown as sticks (PBD 2W1A). B. Modelled *Hpy*CM (orange) and *Hpy*DAH7PS (teal) with potential interface residues shown as sticks.

Complex - In the protein complex structure from *M. tuberculosis* (Figure 4.33 A), the major contribution from *Mtu*DAH7PS to form the *Mtu*CM-*Mtu*DAH7PS interface is from the

inserted α2b helix and its preceding loop, as well as the α7 and α8 helices of the core barrel. *Mtu*CM contributes residues from α1-α2 loop and the α2 and α3 helices. The most notable interactions are contributed by the C-terminus of *Mtu*CM (Gly89 and adjacent residues), which are packed into a shallow groove on the surface of *Mtu*DAH7PS forming interactions with Gln 392 and Arg 399. As for *Hpy*CM and *Hpy*DAH7PS (Figure 4.33 B), although slight interactions between the α1-α2 loop of *Hpy*CM with C-terminus of *Hpy*DAH7PS barrel are possible, severe clashes between the elongated C-terminus (α3) of *Hpy*CM and the α6, α7, α8 helices of *Hpy*DAH7PS are contributed by many charged/polar side chain functionalities causing repulsive interactions at the interface. The differences in the α3 helices of *Hpy*CM and *Mtu*CM, as well as the altered relative position of the *Hpy*DAH7PS barrel due to the twist on the tetramer plane make the complex formation less favourable compared to the *Mtu*CM-*Mtu*DAH7PS system.

In addition, the oligomeric equilibrium in *Hpy*DAH7PS may make complex formation entropically unfavourable as the tetrameric conformation of DAH7PS seems to be a prerequisite. Functionally, *Hpy*CM alone is sufficiently active and displays little or no modifications of its biochemical properties by the presence of *Hpy*DAH7PS and *vice versa*, reinforcing that its interaction with *Hpy*DAH7PS is perhaps not biologically relevant. Nevertheless, the possibility that additional cellular conditions, such as co-factors and ligands, may improve complex formation cannot be ruled out. Future experiments may include screening of optimal conditions and detailed AUC studies for precise detection of the protein complex. Furthermore, *Hpy*CM can be artificially linked to the *Hpy*DAH7PS to minimise the potential entropic disadvantages and encourage more defined functional studies.

Chapter 5. High-resolution methods for studying intramolecular allosteric communication

Preface

Previous chapters explored the details of ligand binding and conformational changes associated with allosteric regulation from several protein systems using combinations of biochemical and physical techniques. While crystallographic data provides important information on a molecular level with atomic resolution, combinations with solution SAXS and other biochemical studies are powerful in illustrating the protein behaviours in biologically relevant states. However, a fundamental problem with those techniques is that important information such as the timescales and dynamics of allosteric motions are often lost due to many experimental limitations. This chapter identifies high-resolution motions from multiple protein conformations derived experimentally, by exploring the feasibilities of applying alternative biophysical techniques including FRET and EPR on available protein systems. This knowledge can be combined with computational methods to further demonstrate the motions associated with allostery.

5.1. FRET

5.1.1. Introduction

Mechanism of energy transfer

The timescale of protein motions is strongly dependent on the nature of the conformational changes (Figure 5.1). Fast motions primarily consist of bond vibrations on femtosecond-picosecond timescales. Side-chain rotations occur within picoseconds. Back-bone rotations of secondary structures are generally in the sub-nanosecond range.[122] Movements or changes of domains are on a slower timescale of microseconds to seconds and are more difficult to examine.[213] Protein motions including local fluctuations involved in side-chain and loop movements and highly corrected large domain conformational changes are of particular relevance to allostery, ranging from picosecond to second timescales.[213]

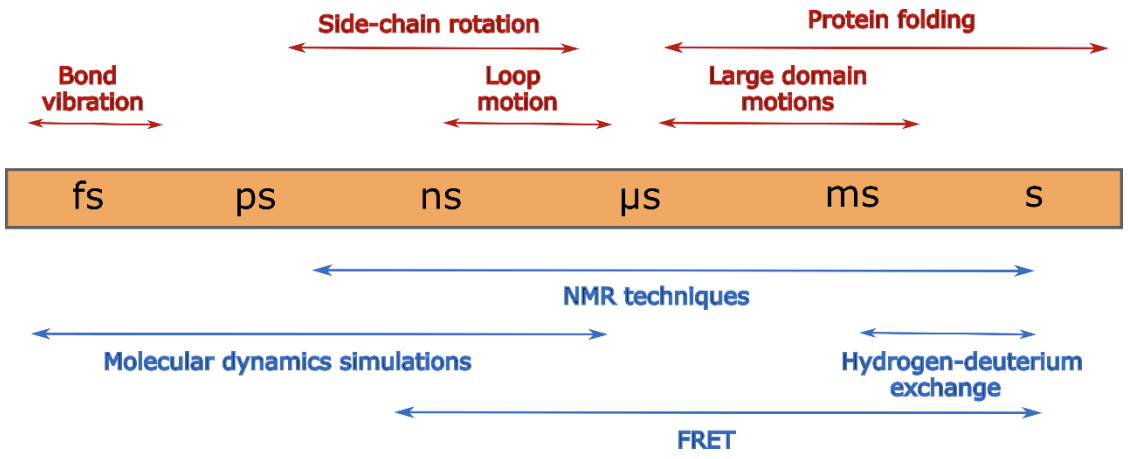


Figure 5.1. Timescale of protein motions and experimental methods often used to study allostery.

Among the techniques often used for studying protein motions associated with allostery, Förster resonance energy transfer (FRET) allows access to motions in the nanosecond to second

timescale. It is an electrodynamic phenomenon when a fluorescence donor molecule in its excited state interacts with an acceptor molecule in its ground state via long-range dipole-dipole forces (Figure 5.2).[214] Factors that are important for FRET efficiency are underlined in the Förster equation (1).

$$(1)\ E = \frac{R_0^6}{R_0^6 + r^6}$$

When the emission spectrum of the donor overlaps with the excitation spectrum of the acceptor, the efficiency of the energy transfer (E) is defined in this equation, where r is the distance between the two fluorophores and R_0 (Förster distance) is the distance at which the energy transfer efficiency is 50%. Förster distance is highly dependent on the quantum yield and relative distance and orientation of the fluorophores as well as the level of spectral overlap. It typically ranges from 20 to 90 Å.[214]

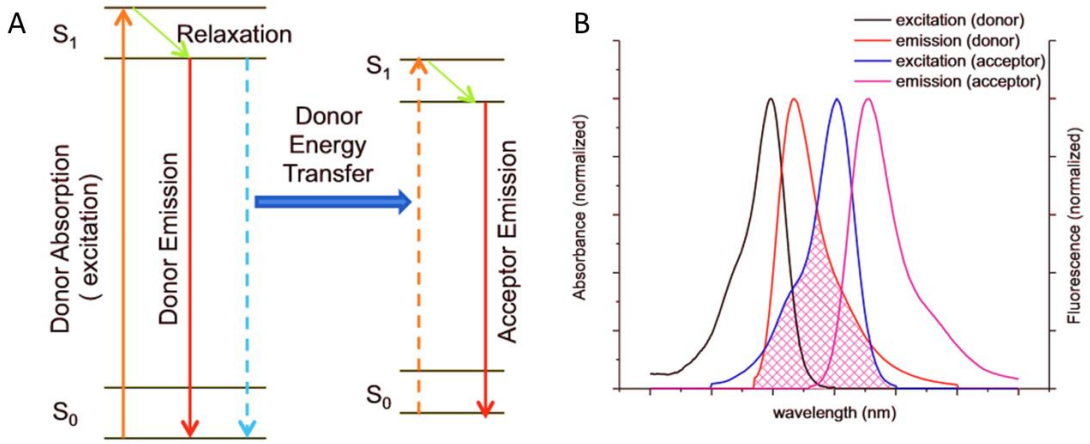


Figure 5.2. Scheme of FRET process. A. Mechanism of FRET. The excited energy of the donor is transferred to the acceptor via FRET. B. Schematic diagram of spectral overlap between donor and acceptor molecules. Figure was adapted from Chen *et al*.[215]

Protein modification strategies

The typical distance range and sensitivity allows FRET to be widely used to study protein interactions,[216] dynamic motions and conformational changes.[214-215, 217] Despite the powerful applications of FRET in the field of chemistry, biology, and medicine in facile, real-time visualisation of a labelled protein, the development of site-specific labelling techniques remains a challenging problem, especially for assessing intramolecular changes.[218]

A few protein modification strategies have been established. For small proteins, short peptide segments can be labelled individually with desired donor or acceptor fluorophores, and combined via chemical ligation.[219] For large proteins, there are four dominant modification strategies.[215] Firstly, a commonly used method is to fuse genetically the protein of interest to a fluorescent protein such as GFP.[220] Although this method is intrinsically specific, the large size and steric bulk of the fluorescent protein pose many problems, including aggregate formation and the disruption of the natural function of the protein of interest.[214, 220] Secondly, pendant enzymes can be fused to label protein using a fluorescent ligand specific to the pendant enzyme, exploiting the enzyme-substrate specificity. Examples of these pendant enzymes include the SNAP-tag,[221] HaloTag,[221] and TMP-tag.[222] They are often smaller than the fluorescent proteins, however, they are still considerably large and often cause steric hindrance problems.[218] Thirdly, modification can be achieved by exploiting a natural or introduced reactive functional group or enzyme recognition site on the protein. This reactive group can then be attached to a fluorescent label via ligation or enzymatic reaction.[215] Enzymes such as microbial transglutaminase[223] and farnesyl transferase[224] are often used to introduce an azide or alkyne group on the protein, which can react with a complementary alkyne- or azide-containing fluorophore via a ‘click’ reaction. This method allows the attachment of small fluorophores with minimal modifications of the protein of interest, although the efficiency of the enzyme-

catalysed reaction and removal of excess free labels can sometimes be problematic.[218] Finally, small fluorogenic probes can react with and attach to minimal peptide tags attached on protein of interest and minimise disruption to the protein. Examples include organoarsenic labels FLAsH and ReAsH.[225-226] This approach achieves specific and undisruptive labelling, but the reagent toxicity and strong background label persistence could be an issue.[227] Besides these dominant strategies, another recent method is the incorporation of unnatural amino acids (UAA) that reacts with the label.[228-229] Although it appears to be an elegant and reliable method for specific labelling, it is currently not generally available due to the difficulties in synthesis and incorporation of UAA, and requirement of specially designed rRNA synthetases.

On the other hand, some non-specific labelling methods are also used. A common method is randomly mixing the protein of interest with both donor and acceptor fluorophores, followed by purification of each labelled product based on natural or engineered properties of the protein.[230-231] Another popular method is step-wise addition of individual labels by controlling the stoichiometry between each dye and the target protein. However, variations in labelling efficiency may severe heterogeneity in the labelled product, which is difficult to address if the mixture cannot be chromatographically separated. These approaches are not selective or site-specific, thus it is difficult to determine the location and types of the fluorophore.

5.1.2. Experiment design

Labelling targets and strategies

Two protein systems were chosen to test the feasibility of applying intramolecular FRET techniques to study conformational changes associated with allostery. These included two Type Iβ DAH7PS enzymes *Tma*DAH7PS and *Gsp*DAH7PS, as both systems have been well studied

and the crystal structures were available for accurate design of labelling sites. After extensive literature research and thorough considerations, enzymatic modification using microbial transglutaminase (mTG) was chosen as the labelling strategy for specific labelling of *Gsp*DAH7PS due to its minimal disruption to the protein and high specificity towards certain substrates (Appendix E, Table S2).

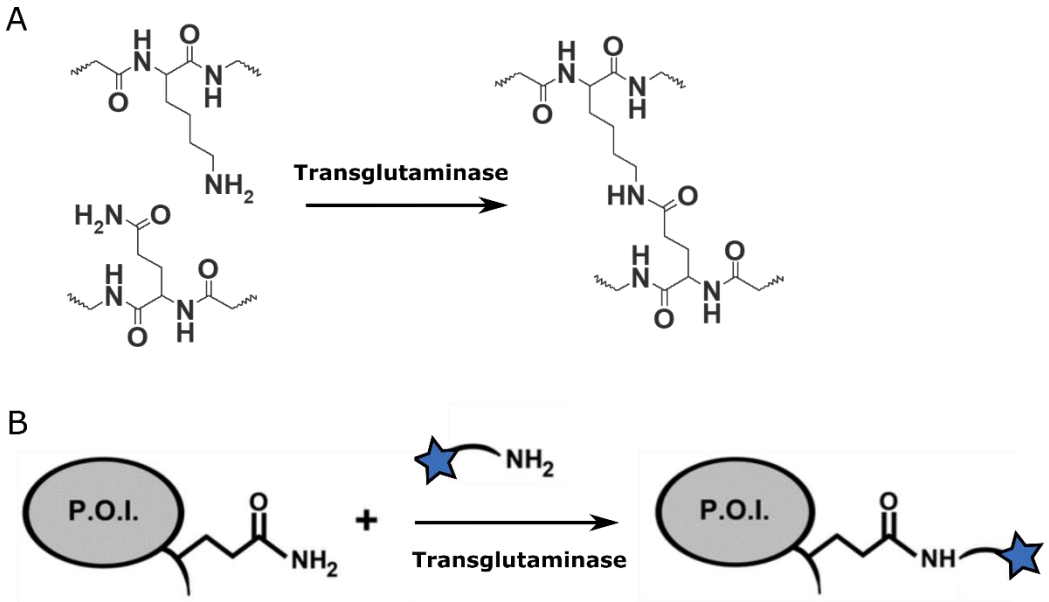


Figure 5.3. Transglutaminase-catalysed reaction (A) and site-specific labelling scheme using TGs (B).

Transglutaminases (TGs) are a family of enzymes that catalyse the acyl transfer between carboxamide groups of Gln and various primary amines, such as the amino group of Lys residue.[232] When a Lys residue serves as an acyl-acceptor, a γ-glutamyl-ε-lysine isopepetide bond is formed between two molecules (Figure 5.3 A). This intrinsic cross-linking property mediated by TGs is important for many biological processes and diseases, such as blood coagulation, type 2 diabetes, and several cancers.[233-235] Microbial TGs are bacterial variants of the TG family. In contrast to other TGs, the mTG-catalysed isopeptide formation is calcium-independent.[236] Generally mTG is expressed as a proenzyme and activated by cleavage of the N-terminus, producing the active enzyme at around 38 kDa, significantly smaller than the

mammalian counterparts (greater than 70 kDa).[236] Due to the broad substrate tolerance of mTGs towards primary amine and high specificity for Gln-containing peptide substrate, these enzymes are attractive catalysts for protein conjugation (Figure 5.3 B).

Collaborations were established with Professor Keillor's research group at the University of Ottawa to explore the application of newly developed high-affinity peptides (Q-tag) as substrates of mTG in labelling of the DAH7PS systems. The Q-tags are short peptides with Gln (Q) recognition site often situated in the centre of the sequence. Hitomi and co-workers reported peptide sequences that could serve as efficient Gln-substrates for mTG by screening of phage-displayed peptide libraries.[237] Subsequently, Keillor's group demonstrated that some selected sequences could be fused to target proteins and discovered three most efficient Q-tags as substrates of mTG, including WALQRPH (Q1), WELQRPY (Q2), and YPMQGWF (Q3).[223, 236] Among these peptides, Q1 was proven to exhibit the highest efficiency to attach to a fluorophore via the mTG-catalysed reaction.[223] Fluorophores with Lys side-chain mimics substitute well for Lys as amine donor. Examples such as cadaverine fluorophores are shown to be successful amine substrates.[238]

Distance estimation and protein modifications

GspDAH7PS - Labelling sites were carefully chosen upon analysis of both open- and closed-form coordinates of *Gsp*DAH7PS. Q-tag WALQRPH was engineered to the N-terminal CM domain of *Gsp*DAH7PS for specific labelling of one fluorophore (Chapter 7). Subsequently, traditional thiol labelling via Cys residue was used to attach the other fluorophore on the DAH7PS catalytic barrel. Thiol reactive probes are frequently used for protein labelling to identify conformational changes, complex formations, and ligand-binding.[239] The primary targets of these probes are Cys residues, which have a low abundance of around 3.3% in proteins.[240] Therefore, in proteins that contain multiple Cys residues, the number of Cys is

usually sufficiently small and easy to be removed by site-directed mutagenesis without significant disruption to the native protein. The low occurrence of Cys also allows saturated labelling with low risk of protein precipitation and fluorophore self-quenching.[240]

The tetrameric nature of *Gsp*DAH7PS posed complications in selection of labelling sites as all possible FRET distances needed to be considered. Residue Asp302 on the DAH7PS barrel appeared to be an ideal site for attaching the second thiol label. In the open form, the distance between the fluorophore pair remained approximately 50 Å, greater than a typical R_0. In the closed form, one pair of fluorophores was located sufficiently close to emit a FRET signal (~28 Å) as shown in Figure 5.4. The difference between the fluorophore distances of the open- and closed-forms was expected to reflect the difference in energy transfer. Therefore, mutation D302C was designed to introduce Cys at position 302. This mutation was unlikely to disrupt catalytic function of the enzyme based on analysis of the crystal structure.

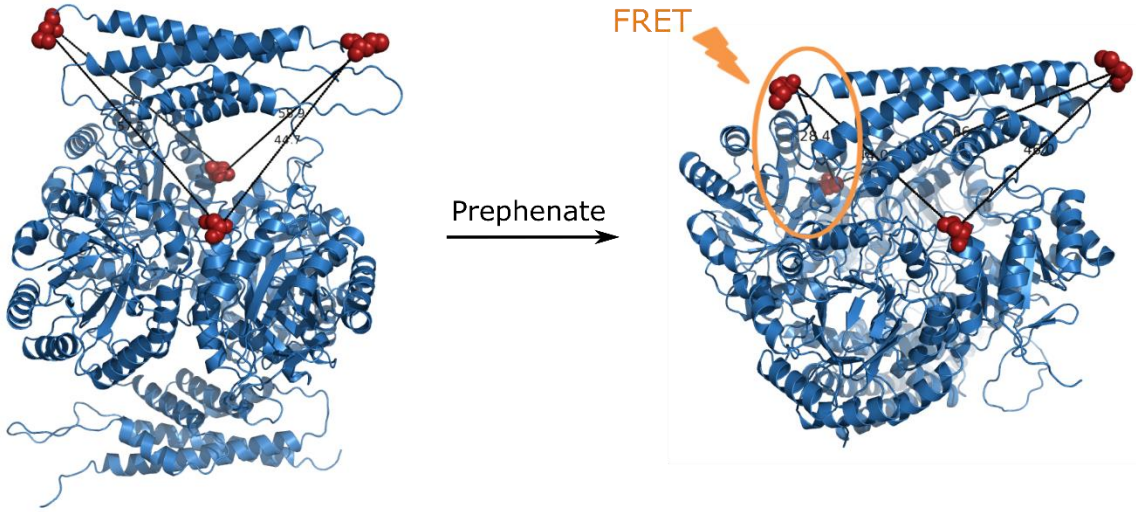


Figure 5.4. FRET design for labelling of *Gsp*DAH7PS. A combination of N-terminus labelling via mTG and Cys-reactive labelling at residue 302 (red spheres) of the barrel was used.

Naturally occurring Cys residues in *Gsp*DAH7PS were also considered for substitution to avoid any unspecific labelling. There are three sequence-encoded Cys residues in *Gsp*DAH7PS, all of which are located on the DAH7PS catalytic barrel (Figure 5.5 A). Both Cys126 and Cys259 appear buried within the barrel, while Cys307 is partially exposed and was replaced by Ser to avoid undesired labelling at this position.

Together, *Gsp*DAH7PS was modified to include the double mutation D302C/C307S and N-terminal modification with Q-tag (*Gsp*DAH7PS$^{D302C-C307S-Q}$). Single mutants including *Gsp*DAH7PS^{D302C}, and *Gsp*DAH7PS^{C307S}, as well as the double mutant *Gsp*DAH7PD$^{D302C-C307S}$ were also created to serve as controls and examine the effect of each modification on protein labelling and function.

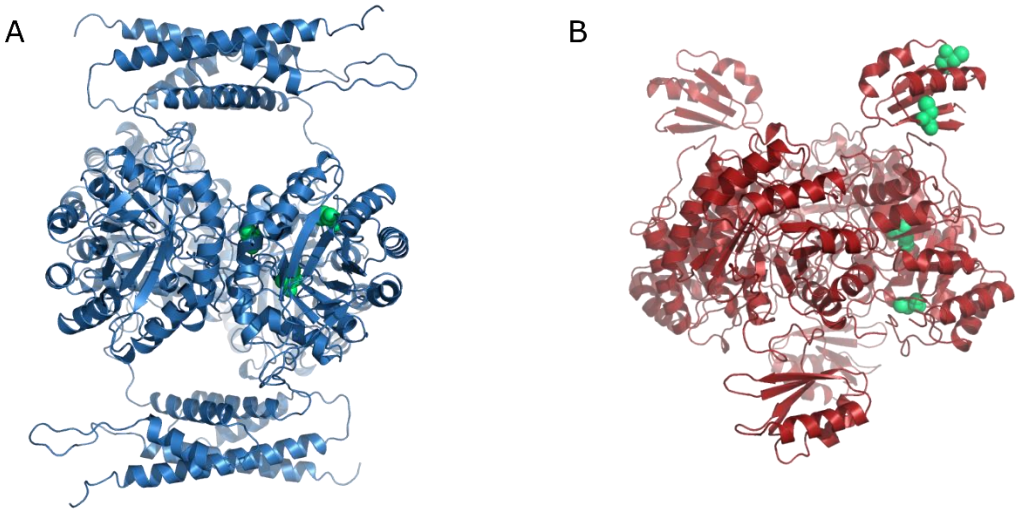


Figure 5.5. Naturally occurring Cys residues in *Gsp*DAH7PS (A) and *Tma*DAH7PS (B). Cys residues from a single chain are highlighted with lime spheres.

TmaDAH7PS - The application of mTG labelling on *Tma*DAH7PS was carefully examined. However, dual labelling with one fluorophore on each domain was not an appropriate approach for this system due to the nature of the ACT domains (Chapter 2). Suitable FRET distances,

which produce sufficient signal differences between the open- and closed-forms, could not be identified. Specifically, the Q-tag could not be attached to the N-terminus of the ACT domain as its movement upon ligand binding is not sufficiently significant for FRET (~13 Å, Figure 5.6 A). C-terminal labelling was not suitable either, as the largest distance displacement was still not sufficient for distinguishing signals between the open- and closed-forms (Figure 5.6 B).

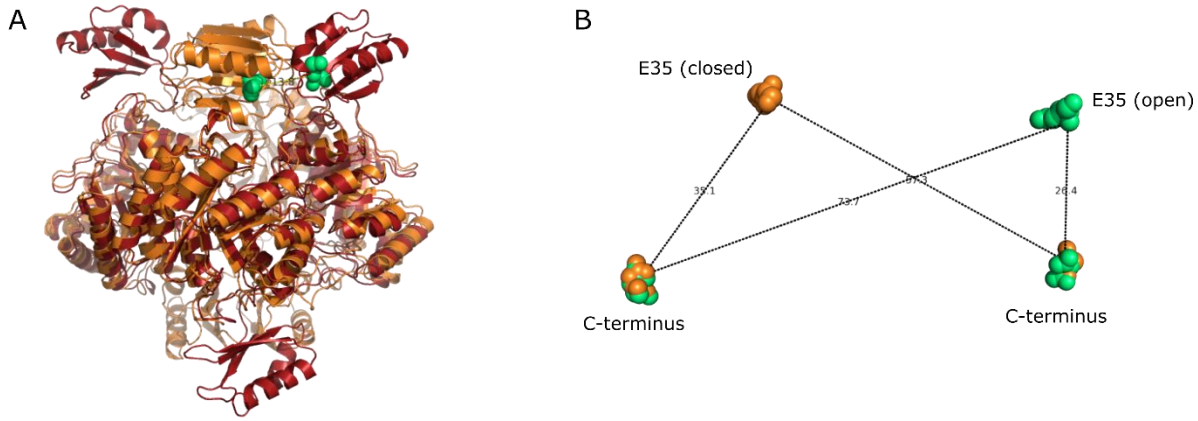


Figure 5.6. Terminal labelling for *Tma*DAH7PS is not suitable for FRET. A. N-terminal labelling with the N-terminus highlighted as lime spheres only produces a distance displacement of around 13 Å between open- and closed-forms. With the engineered Q-tag at this position, this distance displacement would be even smaller, therefore it is not ideal for labelling. B. C-terminal labelling with C-termini from two diagonally opposing chains highlighted as spheres. E35 on the two ACT domains are also highlighted. The open-form highlighted locations are in lime spheres and the closed-from locations are in orange spheres. The highlighted positions in open- and closed-forms would produce similar distances, therefore are not suitable for FRET.

Labelling on solely the ACT domains was used, as the displacement of each ACT domain relative to each other upon ligand binding was sufficient for FRET detection. Because the two

labelling sites were the equivalent sites from each monomeric chain, specific modification of individual site was challenging as it was difficult to introduce different reactive groups on the same residue of different chains. Therefore, a random labelling technique was applied exploiting Cys functionality on the ACT domains with controlled stoichiometry between the number of labelling sites and labels.

There are four naturally occurring Cys residues in each chain of *Tma*DAH7PS (Figure 5.5 B). Two are located in the ACT domain, and another two are located in the DAH7PS barrel. Both Cys28 and Cys58 on the ACT domain are solvent accessible, while residues Cys102 and Cys235 in the catalytic barrel appear buried. Although Cys102 at the active site can be partially exposed transiently, the binding of metal ion and substrate PEP is expected to reduce the flexibility of the catalytic loop and decrease the possibility of label attachments at this position.

Upon inspection of the crystal structures, residue Cys28 on the β2 sheet of the ACT domain provided ideal FRET distances (Figure 5.7). If a pair of thiol-reactive probes were attached to Cys28 residues, the distances between the two probes from diagonally opposing chains would produce the largest displacement of around 60 Å (ligand-free) to 17 Å (with Tyr). On the other hand, residue Cys58 located on the loop region between α3 and β4 of the ACT domain was partially buried based on MD simulations (kindly provided by Dr Eric Lang). Labelling on this residue would produce less ideal FRET distances than Cys28. Therefore, Cys58 of *Tma*DAH7PS was replaced with a Ser residue (*Tma*DAH7PS^{C58S}) to minimise undesirable labelling, and random labelling was designed to target Cys28 with a pair of thiol-reactive fluorophores.

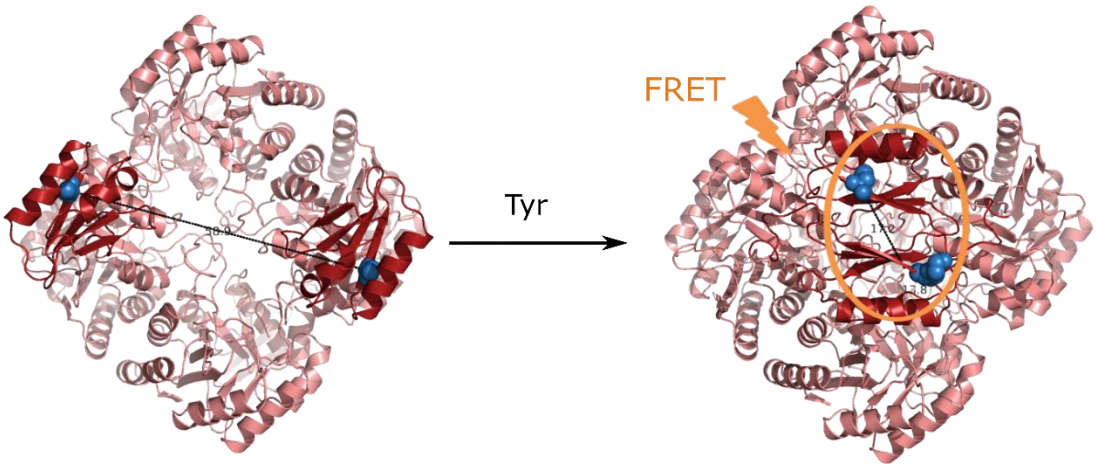


Figure 5.7. FRET design for labelling of *Tma*DAH7PS. Random labelling with controlled stoichiometry at Cys28 (blue spheres) of the ACT domain was used.

It is important to note that the distances estimated based on the open-form crystal structures and models of both *Gsp*DAH7PS and *Tma*DAH7PS are representative of some extreme conformations the proteins can explore. In solution, these distances are expected to vary as the dynamics of the regulatory domains change relative to the positions of each other and of the catalytic domains. Therefore, the FRET distances between the designed labelled sites are expected to sample a range of distances emitting signals with various strength in the absence of ligand, whereas this distance distribution is expected to narrow down as the protein becomes more rigid in the presence of ligand.

Selection of fluorophores

A pair of thiol-reactive fluorophores were required to label Cys28 of *Tma*DAH7PS. A thiol-reactive fluorophore was required to label Cys302 of *Gsp*DAH7PS, and an amine-containing fluorophore was required to serve as substrate of mTG for labelling of the N-terminus of *Gsp*DAH7PS. Based on considerations over the ease of synthesis and the brightness of the

fluorophores, maleimide fluorescein and coumarin were selected for labelling of *Tma*DAH7PS (Figure 5.8).

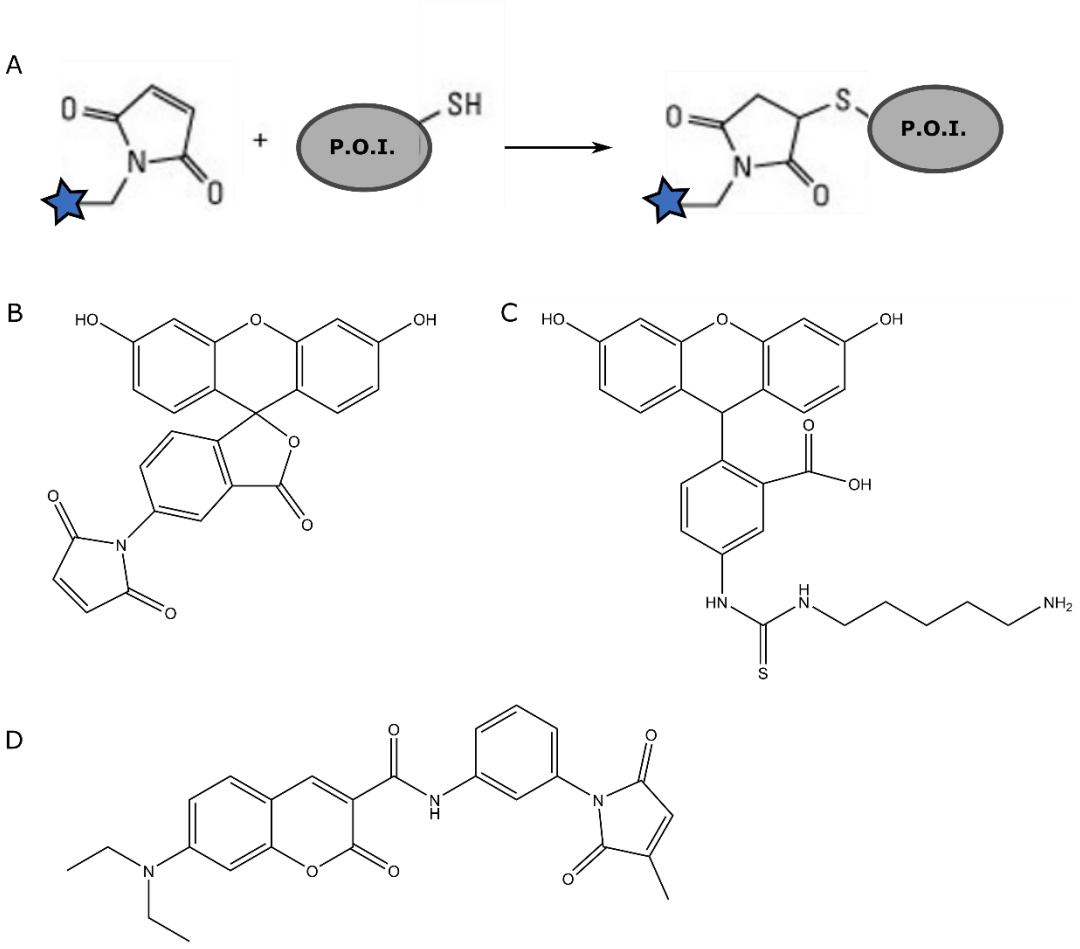


Figure 5.8. Selection of fluorophores. A. Reaction of maleimide probes with Cys side-chain on protein of interest (P.O.I.). B. N-(5-Fluoresceinyl) maleimide used as acceptor for *Tma*DAH7PS labelling. C. Cadaverine form of FITC used as acceptor for *Gsp*DAH7PS labelling. The cadaverine functionality (–NH(CH_2)$_5$NH–) serves as substrate for mTG. D. Diethylamino (methylmaleimide) coumarin used as donor for both *Tma*DAH7PS and *Gsp*DAH7PS labelling.

Maleimides are excellent thiol-reactive reagents for protein labelling and quantitation and do not appear to react with His or Met residues. A thioether bond is formed across the double bond of maleimides in the presence of thiol groups (Figure 5.8 A).[240] The selected N-(5-fluoresceinyl)

maleimide (FITC), purchased from Sigma-Aldrich, is a green fluorescein with excitation and emission maximums of 490 and 510 nm respectively (Figure 5.8 B). The diethylamino (methylmaleimide) coumarin (CPM), kindly provided by Keillor group, is a small blue fluorophore with excitation wavelength of 440 nm and emission wavelength of 485 nm (Figure 5.8 D). The emission spectrum of the donor (CPM) overlaps well with the excitation spectrum of the acceptor (FITC), fulfilling one of the primary conditions for FRET.[215]

A cadaverine form of FITC (FITC-Cd) was synthesised by Keillor group and used as substrate for the mTG-catalysed reaction for labelling of *Gsp*DAH7PS (Figure 5.8 C). It has excitation wavelength of 496 nm and emission wavelength of 519 nm. The K_M value of mTG towards FITC-Cd is around 54 µM in 200 mM MOPS (pH 7.2) and 1 mM EDTA, as per information provided by Keillor group. The donor molecule used was the same CPM molecule for labelling at Cys302. This pair of fluorophores also provides satisfactory spectral overlap.

5.1.3. Mutagenesis and Q-tag engineering

Single-point mutants of *Gsp*DAH7PS^{D302C} and *Gsp*DAH7PS^{C307S} were created using QuikChange® site-directed mutagenesis technique with wild-type *Gsp*DAH7PS plasmid as template (Chapter 7). Double mutation *Gsp*DAH7PS$^{D302C-C307S}$ was produced using the single mutant *Gsp*DAH7PS^{D302C} as template. The Q-tag (WALQRPH) was subsequently fused to the N-terminus of *Gsp*DAH7PS$^{D302C-C307S}$ to produce *Gsp*DAH7PS$^{D302C-C307S-Q}$ using established PCR techniques. For *Tma*DAH7PS, single mutation *Tma*DAH7PS^{C58S} was created via site-directed mutagenesis. All primers used are listed in Chapter 7. All PCR products were treated with DpnI to eliminate contamination from wild-type plasmids. Once DNA sequences were verified to incorporate the correct mutation and Q-tag, the plasmids were transformed,

expressed and purified following the methods for wild-type *Gsp*DAH7PS and *Tma*DAH7PS as described in Chapter 7. Purified products were confirmed to have the correct molecular weights by mass spectrometry (Appendix E, Table S3).

5.1.4. Function of protein variants

Prior to labelling process for FRET, the functions of the protein variants were examined and compared with the parent proteins. Both activity and inhibition assays were performed under the same experimental conditions as the wild-type proteins. All *Gsp*DAH7PS and *Tma*DAH7PS variants were catalytically active with kinetics parameters comparable to that of the wild-type enzymes (Table 5.1, Table 3.1), and more active at elevated temperatures, although the Q-tagged *Gsp*DAH7PS variant appeared slightly less thermostable than other variants. In the presence of prephenate, all *Gsp*DAH7PS variants displayed similar inhibition profiles compared to the wild-type protein (Figure 3.4) with similar values of IC_{50}. In the presence of Tyr, the activity of *Tma*DAH7PS^{C58S} was inhibited by approximately 80%, consistent with the wild-type *Tma*DAH7PS. Therefore, the modifications of the proteins did not significantly disrupt their native functions.

Table 5.1. Kinetic parameters for *Gsp*DAH7PS and *Tma*DAH7PS protein variants.

Protein	DAH7PS activity					CM activity		
	K_M^{PEP} (μM)	K_M^{E4P} (μM)	k_{cat} (s^{-1})	k_{cat}/K_M^{PEP} ($\mu M^{-1}s^{-1}$)	k_{cat}/K_M^{E4P} ($\mu M^{-1}s^{-1}$)	$K_M^{chorismate}$ (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)
*Gsp*DAH7PS^{D302C}	44 ± 2	49 ± 4	30 ± 1	0.68	0.61	101 ± 9	4 ± 0.1	0.04
*Gsp*DAH7PS^{C307S}	55 ± 3.6	60 ± 6	32 ± 0.8	0.58	0.53	100 ± 8.6	3.2 ± 0.2	0.03
*Gsp*DAH7PS$^{D302C\text{-}C307S}$	48 ± 4	58 ± 3	29 ± 1.5	0.60	0.5	95 ± 7	3.5 ± 0.1	0.04
*Gsp*DAH7PS$^{D302C\text{-}C307S\text{-}Q}$	50 ± 3	61 ± 7	33 ± 2.5	0.66	0.54	98 ± 5	3.0 ± 0.3	0.03
*Tma*DAH7PS^{C58S}	11 ± 1	18 ± 2	17 ± 2	1.55	0.94	NA	NA	NA

5.1.5. Preparation of mTG

Gene expression

Plasmid containing the gene encoding for the *Streptomyces mobaraensis* mTG was kindly donated by Professor Markus Pietzsch from Martin Luther University of Halle-Wittenberg. The plasmid pDJ1-3 encodes for mTG with the N-terminal proenzyme sequence and a C-terminal His_6 tag. It was transformed into *E. coli* BL21*(DE3) cells by using standard procedures described in Chapter 7. An autoinducing medium was used for expression of mTG. A small preculture (~5-10 mL) was incubated overnight at 37 °C with shaking at 220 rmp in ZYP-0.8G medium (Chapter 7).[241] The preculture was then used to inoculate a large culture of around 500 mL in ZYP-505 medium,[241] which was subsequently incubated at 37 °C for two hours followed by overnight incubation at 22 °C with shaking at 220 rpm. The cells were harvested following standard procedures. Cell lysis was performed using standard on-ice sonication methods in 0.2 M Tris-Cl buffer at pH 6.0. Crude lysate was centrifuged at 14,000 g for 30 min at 4 °C.

Purification

Activation of mTG by cleavage of the N-terminal pro-sequence was then performed by adding trypsin to the crude soluble fraction with ratio of 1:9 (v/v). The mixture was incubated at 30 °C for 45 minutes. The activated mTG was purified by using standard IMAC procedures exploiting the C-terminal His_6 tag (Figure 5.9). The IMAC column was equilibrated in 50 mM phosphate buffer (pH 8.0) containing 300 mM NaCl. The desired protein was eluted with 0 to 140 mM imidazole gradient. A final step using SEC (0.2 M Tris-Cl, pH 6.0) was optional depending on the level of purity. Typical yield of activated mTG was approximately 80 mg per litre of growth.

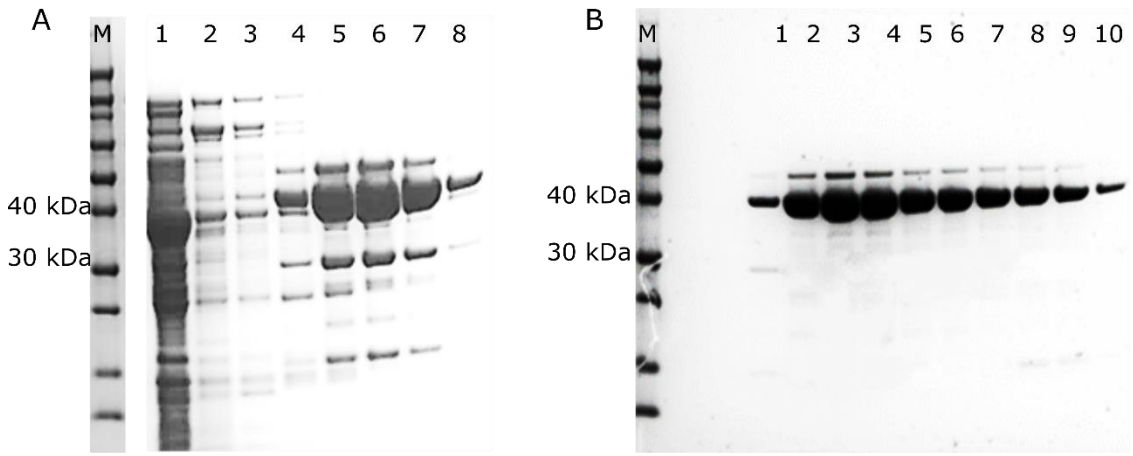


Figure 5.9. Purification of pro-mTG. A. First step purification using IMAC column. M, marker. 1, flow-through fraction. 2-3, washed fractions. 4-8, eluted fractions containing mTG. B. Second step using SEC column. 1-10, eluted fractions containing purified mTG.

Activity test

The catalytic activity of mTG was tested by applying the hydroxamate assay described by Folk and Cole.[242] Purified mTG was incubated with substrates 100 mM hydroxamate and 30 mM Z-Gln-Gly (purchased from Sigma-Aldrich) in SEC buffer at 37 °C for 10 minutes. The reaction was then quenched by mixing with a solution containing 2 M $FeCl_3$, 300 mM trichloroacetic acid, and 800 mM HCl. The iron complex produced was easily detected by eye and its absorbance could be measured at 535 nm after precipitate was removed by centrifugation (Appendix E, Figure S12). Typical activity of mTG was 1 μmol product formation (glutamic acid γ-hydroxamate) per unit (U) of mTG per minute at 37 °C.

5.1.6. Labelling of *Gsp*DAH7PS

FITC-cadaverine labelling

The activated mTG was used to attach FITC-Cd on the N-terminus of *Gsp*DAH7PS$^{D302C-C307S-Q}$. Initially experimental procedures followed methods established by Oteng-Pabi *et al.*[223] The purified protein (final concentration 0.5 mg·mL^{-1}) was added to a solution containing 500 µM FITC-Cd (approximately ten-times of the K_M value), 200 mM MOPS (pH 7.2), and 1 mM EDTA. One unit (U) of mTG was added to the mixture to initiate the transamidation reaction at a total volume of 500 µL. The reaction was then incubated at 37 °C for two hours with gentle stirring. Small aliquots were taken periodically at time 0, 15, 30, 60 and 120 minutes to record the progress of labelling.

Visible protein precipitates were observed during the labelling reaction. The labelled product was centrifuged and the soluble fraction was applied on a PD-10 column and subsequently washed repeatedly over a membrane with 10 kDa cut-off to remove free labelling reagents. The effect of labelling was assessed by mass spectrometry and the resulting molecular weight of the product was consistent with the unlabelled protein. Labelling conditions were then varied in attempt to improve the result. A range of alternative reaction buffers were tested including BTP, phosphate, and Tris-Cl to reduce precipitation. However, severe level of precipitation was still observed with different buffers. Various reaction time, temperatures, substrate (FITC-Cd) and protein concentrations, as well as higher FITC-Cd to protein ratios were also tested to determine whether the protein that remained in the soluble fraction could be labelled efficiently. However, these attempts did not successfully produce soluble labelled protein. Additionally, the DAH7PS activity could not be detected after the labelling reaction.

A negative control with the same labelling conditions was performed except that mTG was not included (i.e. target protein and free label only), therefore no conjugation of the FITC-Cd probe

was expected. After incubation at 37 °C for two hours, the protein was tested to be catalytically active and remained unlabelled as verified by MS, suggesting that the presence of mTG or the mTG-catalysed reaction posed disruption to the native function of the protein. Another control was performed using the *Gsp*DAH7PS$^{D302C-C307S}$ mutant incubated with FITC-Cd and mTG protein under the same conditions. After incubation, this protein also remained catalytically active and unlabelled. This control verified that it was likely the interaction of mTG with the Q-tag attached on the *Gsp*DAH7PS$^{D302C-C307S-Q}$ variant that destabilised the target protein, rather than the presence of mTG causing random interactions with the target protein. This control verified that the naturally occurring Gln (Q) residues in *Gsp*DAH7PS were unlikely to react as substrates for mTG, and the only point of interaction appeared to be the engineered Q-tag, demonstrating the sufficient specificity of this system.

Dansylcadaverine labelling

The initial mTG labelling experiments showed that attachment of FITC-Cd on *Gsp*DAH7PS perturbed the protein function. This was perhaps due to the size and hydrophobicity of the probe. Therefore, an alternative probe dansylcadavarine (DNS-Cd) was selected and tested to determine if it was less obtrusive than FITC-Cd (Figure 5.10). Dansylcadaverine is a smaller labelling reagent (Mw 335.46 g·mol^{-1}) with the same cadaverine functionality (–NH$(CH_2)_5$NH–) that serves as substrate for mTG. It has excitation wavelength of 335 nm and emission wavelength of 512 nm.

Similar labelling procedures were applied for the mTG-catalysed reaction between DNS-Cd and *Gsp*DAH7PS$^{D302C-C307S-Q}$ as previously described for FITC-Cd labelling. After around 20 minutes of incubation at room temperature, some white aggregates were observed, however less severe compared to the FITC-Cd labelling. A diluted reaction (2 mL instead of 0.5 mL) was used to reduce precipitation. After two-hours incubation at room temperature, the

remaining soluble fraction was isolated and washed (Appendix E, Figure S13). Although this fraction showed weak DAH7PS activity, MS results showed that the soluble protein remained unlabelled even with varied labelling conditions such as altering reaction temperatures and label concentrations.

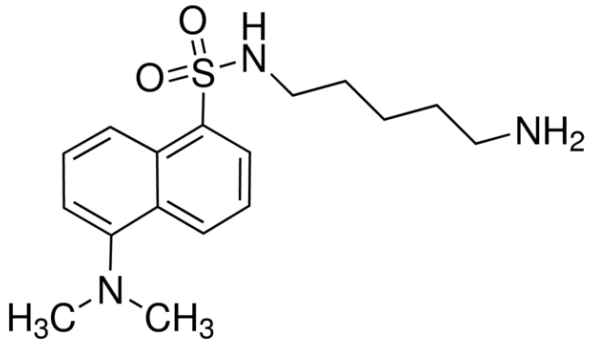


Figure 5.10. An alternative probe for mTG-mediated *Gsp*DAH7PS labelling. Dansylcadaverine is a broadly applied small amine probe that serves as substrate for mTG.

Overall, the mTG-mediated labelling at the N-terminus of *Gsp*DAH7PS was troublesome and disruptive to the native protein folding and function. Both FITC-Cd and DNS-Cd labelling were unsuccessful in producing soluble, functional, and labelled protein. This suggests the modifications at the extended N-terminal loop region disturbs the protein significantly, although this region is located remotely from the DAH7PS active site. Fluorophore labelling in combination with the extended Q-tag possibly affects the flexibility of the loop, which may play an important role in the stability and activity of the *Gsp*DAH7PS. Furthermore, the addition of hydrophobic label may also destabilise the protein.

5.1.7. Labelling of *Tma*DAH7PS

Reaction setup

Previously described random labelling strategy (Chapter 5.1.2) for *Tma*DAH7PS^{C58S} using controlled stoichiometry of FITC and CPM was performed following standard methods for thiol-reactive probe labelling with some modifications.[240] There are four Cys targets per molecule of *Tma*DAH7PS^{C58S} (one Cys per chain), therefore two CPM (donor) molecules and two FITC (acceptor) were mixed with one molecule of *Tma*DAH7PS^{C58S}, which was the equivalent of molar ratio 2:1:1 of the target protein, CPM and FITC. Protein stock was treated with ten-fold molar excess of dithiothreitol (DTT) for reduction of disulphide bonds. Thiol modification was then performed under inert atmosphere (N_2) to minimise thiol oxidation. Approximately 57 µM target protein was prepared in 50 mM BTP buffer (pH 7.5) at room temperature. A 1:1 FITC and CPM (final 28.5 µM each) solution was freshly prepared prior to addition into the protein solution dropwise over a half-hour period. All thiol modification steps were carried out in the dark and the reaction was incubated at room temperature for two hours with gentle stirring. Upon completion of the reaction, the mixture was passed through a PD-10 column to remove the remaining free labels from the protein. A fast-moving fluorescent band containing labelled protein was observed under natural light and the corresponding fractions were verified by NanoDrop^{TM} and collected. Free labels were eluted with subsequent buffer washes. An additional washing step of the labelled protein over a 10-kDa membrane was sometimes included to remove any remaining labels from the protein (Figure 5.11).

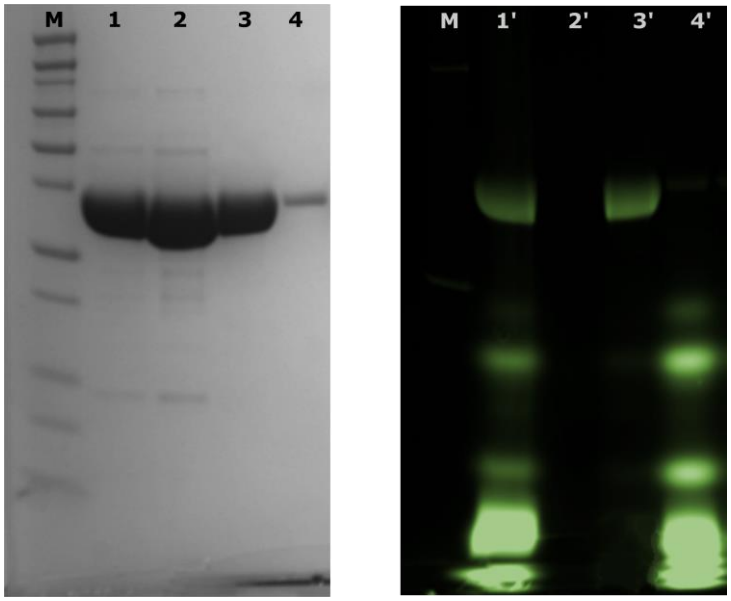


Figure 5.11. FITC and CPM labelled *Tma*DAH7PS^{C58S} under white light (left) and fluorescence (right). M, marker. 1 (1'), crude labelled product. 2 (2'), unlabelled control. 3 (3'), purified and washed labelled product. 4 (4'), removed impurities and free labels.

The size of the labelled product was analysed by MS. The results showed that majority of the protein was successfully labelled with FITC, however, very small amount had CPM attached, suggesting different labelling efficiencies of the two labels towards the same Cys28 residue. Therefore, the relative concentration of CPM to FITC was increased to enhance the attachment of CPM on Cys28. As some precipitation was observed during the thiol modification reaction with the previous experimental setup, the protein concentration was also reduced from 57 to 28.5 µM. The new setup included 28.5 µM target protein, 14.25 µM FITC and 42.75 µM CPM (2:1:6 ratio of protein:FITC:CPM) with elongated incubation time of an additional eight hours at 4 °C. Although more CPM-labelled protein was easily detected by MS, the FITC-labelled protein was still significantly more dominant (~70% of labelled protein). The CPM concentration was then further increased to up to 15 M excess of the protein, however its labelling efficiency remained poor compared to FITC.

To avoid the competitive labelling of CPM and FITC and improve the CPM labelling, CPM was added prior to addition of FITC, instead of being added simultaneously. Final concentrations of 28.5 µM protein and 14.25 µM of each label were applied (2:1:1). CPM was added slowly into the protein solution first. The reaction was incubated at 4 °C overnight before the FITC aliquot was added and allowed to react for two hours at room temperature. This method significantly improved the production of CPM-labelled protein and further tests determined the optimal molar ratio of CPM and FITC was 3:1 with the stepwise reaction scheme to produce around equal amount of CPM- and FITC-labelled proteins. While this procedure achieved the ideal labelling result, some unlabelled protein was detected by MS. Therefore, the concentrations of both labels relative to the protein were assessed and increased accordingly. The final recipe contained molar ratio of 1:1:3 of protein:FITC:CPM with the stepwise procedure.

Signal detection

Fluorescence intensities were measured by using a Varian Cary Eclipse Fluorescence Spectrophotometer. Initial scanning of the BTP buffer showed interference at 384 nm excitation, therefore alternative buffers including HEPES, MES, MOPS and phosphate were examined. The labelled *Tma*DAH7PC^{C58S} remained catalytically active in HEPES, MES and MOPS buffers and became inactive in phosphate buffer. HEPES buffer displayed high background absorbance at 232 nm. Therefore, MOPS buffer was chosen for subsequent measurements.

The fluorescence intensities of the labelled *Tma*DAH7PS^{C58S} were measured in the absence or presence of Tyr (Figure 5.12). Upon addition of Tyr, the protein was expected to undergo conformational change from the open to closed form as previously described, emitting FRET signal in which the acceptor intensity should increase. However, the addition of Tyr did not

result in obvious increase in the relative intensity of the acceptor (emission λ = 517 nm) compared to the apo protein.

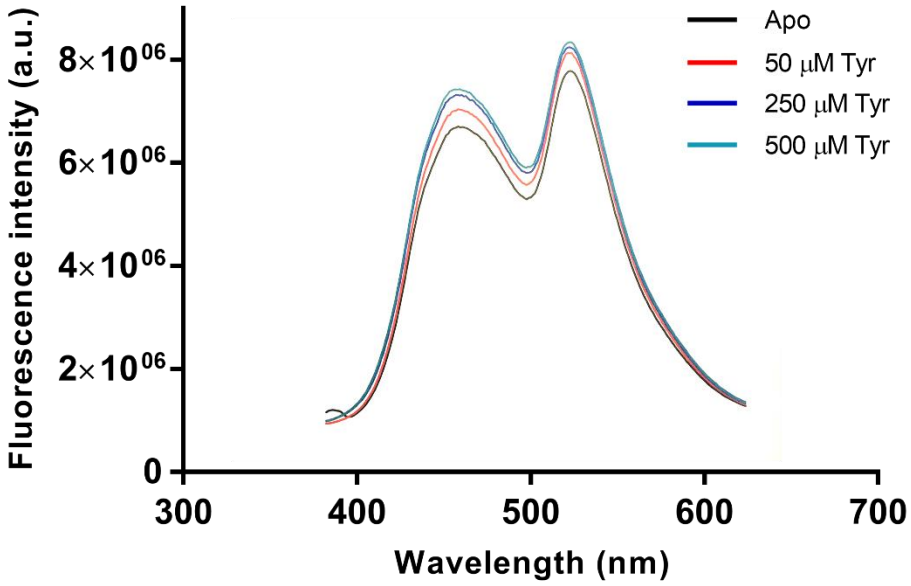


Figure 5.12. Fluorescence spectra of the dual labelled *Tma*DAH7PS^{C58S} at different concentrations of Tyr. Intensities were recorded at room temperature with approximately 2 x 10^{-7} μM protein in 10 mM MOPS buffer (pH 7.5).

Due to the tetrameric nature of the protein, the non-specific labelling strategy used for *Tma*DAH7PS^{C58S} would have resultd in a heterogenous population of species, including molecules labelled with four donors (D), four acceptors (A), three Ds and one A, three As and one D, and two Ds and two As. Due to the small differences in the size of these species, they could not be separated effectively. This heterogeneity posed high background intensity as only proteins with dual labels would produce desired intermolecular energy transfer. The slight decrease in donor and/or increase in acceptor emission from intramolecular FRET may not be detectable with high background from non-specific intermolecular interactions.

5.2. EPR

5.2.1. Introduction

The function of proteins is determined by their structure, therefore precise structural information on all conformations accessible to the proteins is crucial in understanding their function. Biophysical techniques such as X-ray crystallography and NMR are powerful in providing high-resolution atomic structures of proteins. However, X-ray crystallography relies on successful production of protein crystals and provides only snapshots of the protein structure that represents one of the low-energy states under crystal constraints. While NMR can be used to obtain protein structures in solution, it becomes problematic for large proteins exceeding around 50 kDa.[243] As an alternative, electron paramagnetic resonance (EPR) spectroscopy has been widely employed to provide structural information with atomic resolution on a local scale for proteins in solution or solid state without size restrictions. Distance information determined by EPR relies on magnetic dipole-dipole interaction between magnetic moments of two spin probes often attached on desired sites of the target protein.[244]

Pulsed electron–electron double resonance (PELDOR), sometimes referred to as double electron–electron resonance (DEER), is a powerful pulsed EPR technique capable of detecting long range spins of up to 80 Å, making it complementary to FRET.[243] While FRET can provide real-time dynamic information on single-molecule level in solution, PELDOR typically requires frozen samples at micromolar concentration range. PELDOR delivers accurate distance measurements within 1 Å precision and distribution ensembles reflecting accessible conformational dynamics of the protein.[243] This feature is particularly powerful in studying dynamic systems, such as *Tma*DAH7PS. Instead of a static snapshot of one crystallographic conformation, a range of dynamic conformations can be sampled so that the real conformational dynamics associated with the ACT domain can be accessed. With the long-

range distance information, more accurate biomolecule structures can be derived in conjunction with complementary techniques such as X-ray crystallography and molecular dynamics simulations.[245-246]

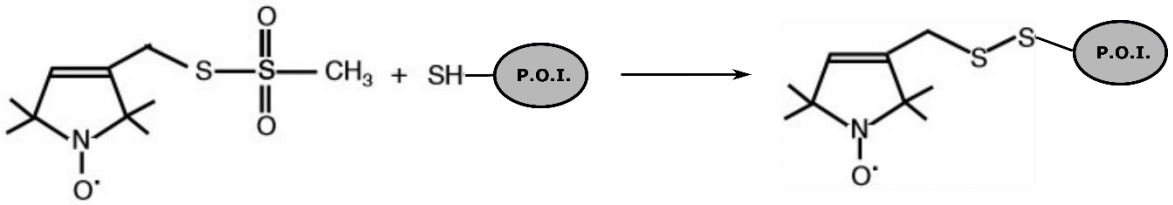


Figure 5.13. Attachment of MTSSL spin label on Cys residue on protein of interest via disulphide bond formation. The PELDOR signal arises from the unpaired electron on the nitroxide group.

As with other EPR methods, distance measurements from PELDOR experiments rely on the dipolar interaction between two paramagnetic probes. Types of these paramagnetic centres often include metal ions, clusters, amino acid radicals, organic radical cofactors, or artificially engineered spin labels.[247] Most PELDOR experiments are based on site-directed spin labelling (SDSL). The most commonly used SDSL technique for proteins involves labelling of the target sites with (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)-methanethiosulfonate (MTSSL) spin label.[248] MTSSL reacts specifically with the thiol functionality of Cys residues and forms disulphide bonds (Figure 5.13).[248] Cys residues can be introduced to the desired sites on proteins by mutagenesis and subsequently be attached to MTSSL labels. Advantages of applying SDSL with MTSSL label include its high selectivity towards thiol groups and it allows some degree of flexibility at labelling sites and minimal perturbation of the protein.[247]

5.2.2. Experiment design

As described in previous chapters, the unique ACT domain-mediated allostery in *Tma*DAH7PS and *Tye*DAH7PS has been examined using combinations of techniques such as X-ray crystallography and SAXS. While the crystal structures of two extreme conformations representing the open and closed forms of *Tma*DAH7PS are obtained in the absence or presence of ligand Tyr respectively, and SAXS experiments reveal the solution conformational change associated with allostery, the real dynamics of the ACT domain and its accessible conformation distribution remain unknown.

The previously designed *Tma*DAH7PS^{C58S} mutant (Chapter 5.1) was selected with Cys28 of the ACT domain as labelling target for MTSSL. MTSSL attachment on Cys28 provides ideal distance for PELDOR measurements (Figure 5.7) and previous FRET experiments showed that successful thiol-reactive labelling was achieved via Cys28 and resulting labelled protein remained functional.

5.2.3. Spin labelling

The MTSSL spin labelling procedures were similar to that of the FRET labelling described in Chapter 5.1.7. Purified *Tma*DAH7PS^{C58S} was treated with 10 mM DTT at 4 °C overnight to ensure complete reduction of Cys residues and effective subsequent disulphide bond formation with the spin label. Upon removal of DTT, MTSSL stock (1 mg·mL^{-1} in DMF, purchased from Toronto Research Chemicals) was added in the protein solution dropwise over a half-hour period with gentle stirring at room temperature under inert atmosphere. The final concentration of protein was 100 μM, and concentration of MTSSL was 1 mM in 50 mM BTP (pH 7.5). The

reaction was incubated at room temperature for four hours and at 4 °C overnight before the free label was washed and removed the next day.

MS results indicated that significant portion of the product had a molecular weight consistent with the target protein attached to two MTSSL labels per chain. This was unexpected as only a single label was attached on the same protein during FRET labelling. Molar ratio of MTSSL label to protein was then reduced from 10:1 to 1:1 to limit MTSSL labelling on two sites of the protein. However, the labelled product still contained a substantial amount of dual labelled protein. Reaction time was then reduced to two hours at room temperature followed by six-hours incubation at 4 °C. However, most protein remained unlabelled with the reduced reaction time.

There are three Cys residues in *Tma*DAH7PS^{C58S}, including the target labelling site at Cys28, active site Cys102 and Cys235 in the catalytic barrel (Figure 5.5 B). In contrast to the buried residue Cys235 (accessible area 0 Å^2) in the DAH7PS barrels, Cys102 appears partially exposed transiently with accessible area of 32.29 Å^2 therefore is more likely to be labelled due to the small size of MTSSL compared to FRET labels.[132] Residue Cys102 is located in the active site of the DAH7PS catalytic barrel and responsible for interaction with metal ion essential for catalysis. Upon binding of substrates and metal ion, the accessible area of Cys102 is reduced to 5.25 Å^2 (Appendix E, Figure S14). Further test on the activity of the MTSSL labelled *Tma*DAH7PS^{C58S} revealed that while the labelled product remained catalytically active and showed inhibition in the presence of Tyr, the specific activity was reduced to approximately 60% compared to that of the unlabelled protein.

Docking experiments conducted by Dr Wanting Jiao suggest that the MTSSL probe adopts two possible conformations in or out of the active site pocket (Figure 5.14). In Conformation 1

(Figure 5.14 A), MTSSL occupies the active site, substrates PEP and E4P can no longer bind to this pocket. As the active sites are generally optimised for substrate binding, it is possible that binding of PEP and E4P may displace MTSSL and force it outward into the alternative conformation with rotated Cys side-chain (Conformation 2, Figure 5.14 B). In Conformation 2, MTSSL can be stabilised by similar degree of interactions with the nearby loop compared to the interactions of Conformation 1 with the active site. Coordination of divalent metal ion with Conformation 2 is possible through alternative interactions allowing catalysis to occur. The predicted MTSSL binding conformations and the heterogeneity of the labelled product (mixture between single and dual labelled proteins) support the reduced activity of the protein upon labelling.

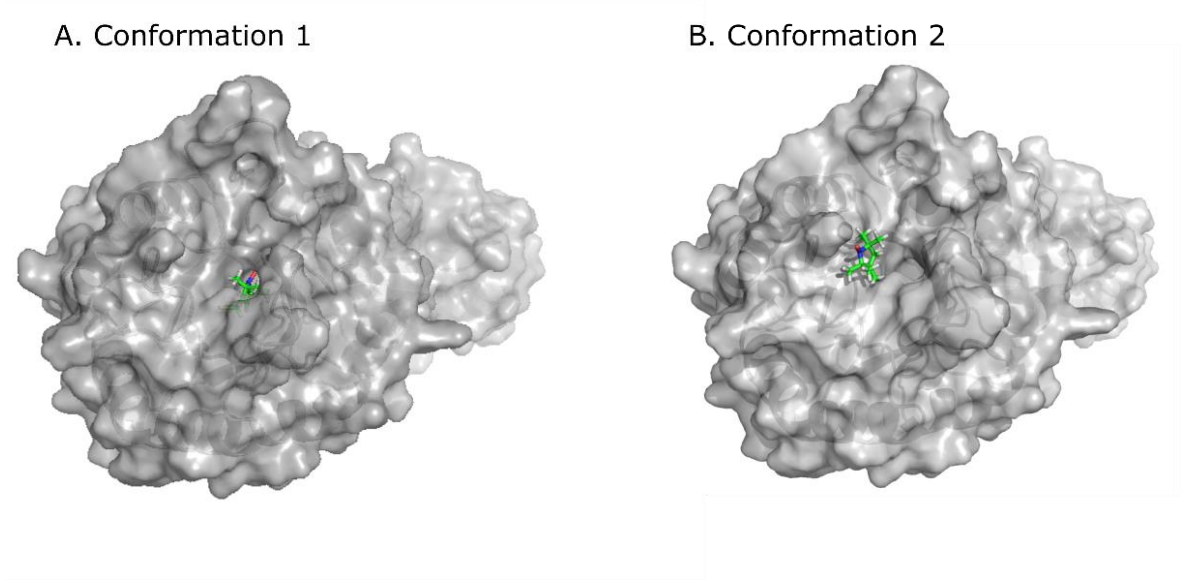


Figure 5.14. MTSSL binding conformations on Cys102 of *Tma*DAH7PS^{C58S}. Docking experiments reveal 2 possible conformations of MTSSL (green sticks) in or out of the active site pocket when attached to Cys102. Protein is shown as grey surface.

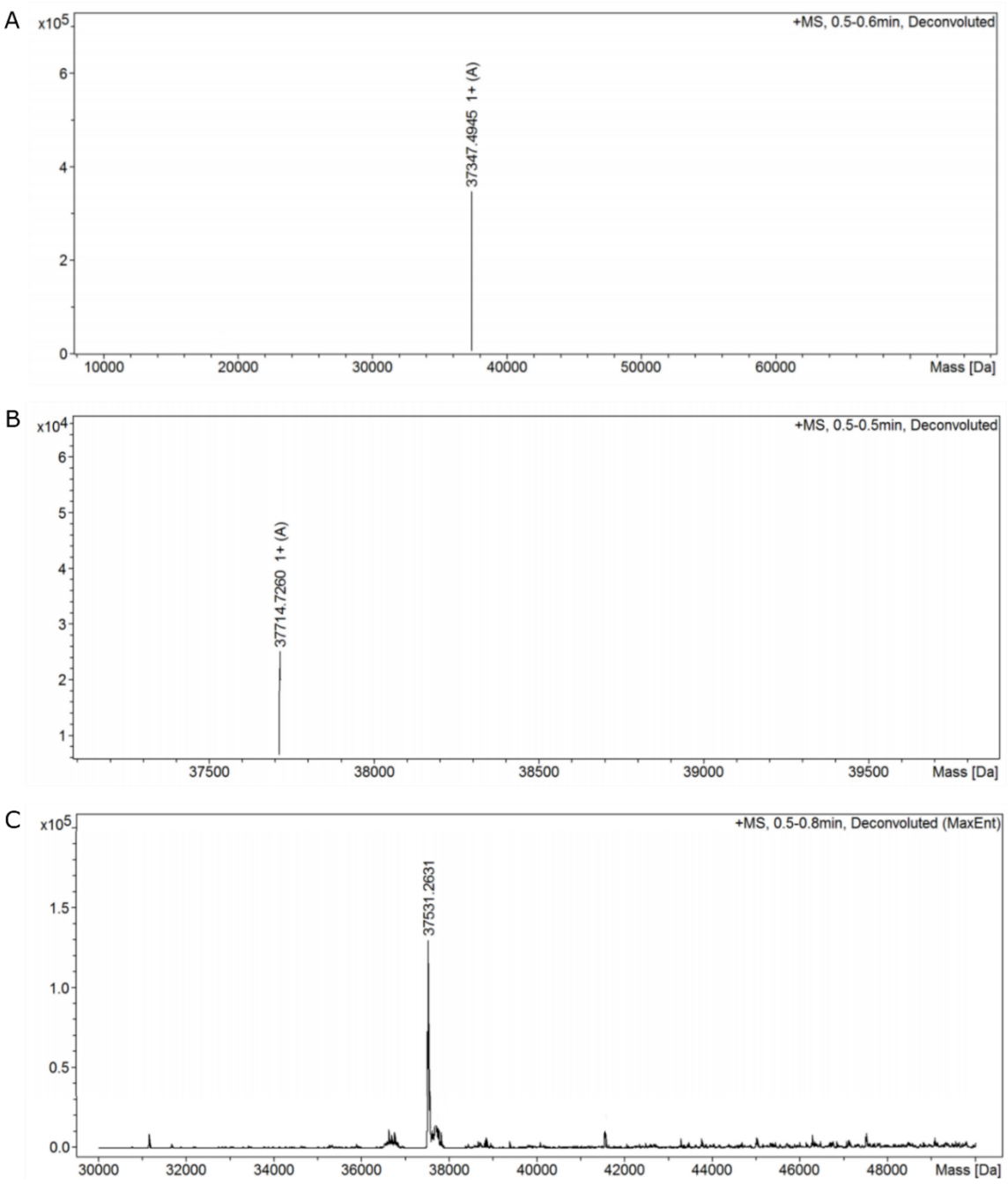


Figure 5.15. MS results for MTSSL labelling of *Tma*DAH7PS. A. Unlabelled control had Mw of 37347.49 Da, consistent with theoretical Mw of 37346.04 Da. B. In the absence of Mn^{2+} and PEP, the protein was labelled with 2 MTSSL molecules with Mw of 37714.73 Da, consistent with the calculated value of 37716.6 Da. C. With the presence of 1 mM Mn^{2+} and 1 mM PEP in the labelling reaction, the target protein was labelled with only one MTSSL and had Mw of 37531.26 Da, consistent with the theoretical value of 37531.3 Da.

Conditions for the labelling reaction was therefore modified based on the analysis above to produce target protein with a single MTSSL label per chain. A solution containing 1 mM Mn^{2+} and 1 mM PEP was added into the reaction to protect the Cys102 from reacting with MTSSL with other experimental conditions remained unchanged. This approach was proven successful in producing single labelled product as verified by MS (Figure 5.15). The activity of the labelled protein was tested and it retained normal activity and showed inhibition in the presence of Tyr. As a control, a C235S mutant was also produced using *Tma*DAH7PS^{C58S} as template to generate the *Tma*DAH7PS$^{C58S\text{-}C235S}$ double mutant. When the same labelling conditions were applied using this double mutant, two MTSSL label were attached in the absence of Mn^{2+} and PEP. Only a single label was identified when Mn^{2+} and PEP were present in the labelling reaction, as was observed for the *Tma*DAH7PS^{C58S} mutant. This control supported the prediction that the Cys102 residue was labelled with MTSSL in the absence of the additives, rather than the buried Cys235; and that Cys102 could be protected by adding Mn^{2+} and PEP in the labelling reaction.

5.2.4. MMM analysis

PELDOR distance measurements can be interpreted appropriately only if the possible conformations of the spin label are considered.[249] Due to the level of freedom around the five rotatable bonds between the electron spin and the Cys protein backbone (Figure 5.16), the distance distributions measured by PELDOR may vary.[247, 249] Therefore, all possible rotamers of MTSSL attached on the *Tma*DAH7PS must be analysed for accurate interpretation of the experimental data. Multiscale modelling of macromolecules (MMM) software is for modelling, visualisation and analysis of models of spin labels on target protein using the rotamer library approach based on restrains from several experimental techniques.[250-251]

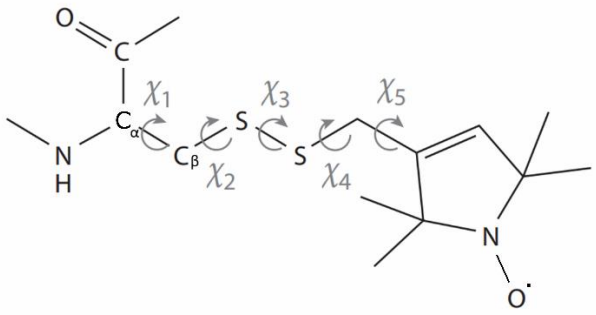


Figure 5.16. Rotatable bonds of MTSSL-Cys conjugation. This freedom causes variations in the distance distributions of the electron spin with respect to the C_α atom of the Cys residue in the protein. Dihedral angles χ_1-χ_5 are indicated.

MMM analysis for *Tma*DAH7PS system was performed with kind assistance from our collaborator Dr Alistair Fielding (University of Manchester). Coordinates from crystal structures PDB 1RZM and 3PG9 of *Tma*DAH7PS were used as input for the analysis. Figure 5.17 shows the distance distributions between the MTSSL spin labels attached to Cys28 derived from the top 26 highest populated rotamers. In the open form, Cys28 residues on the β2 sheets of the opposite ACT domains were well separated with an average distance of around 8 nm. The spin location at the Cys28 residues were also relatively close to the equivalent sites from adjacent subunits (A-G and B-F, Figure 5.17 A), producing peaks at around 6 nm. Other possible pairwise contributions produced interspin distances at approximately 7 nm and 8 nm. In the closed form, labelled sites became close together (A-F and B-G, Figure 5.17 B) producing peak in distance distribution at around 2 nm. Other contributions from possible pairwise distances yielded peaks at around 8.5 nm.

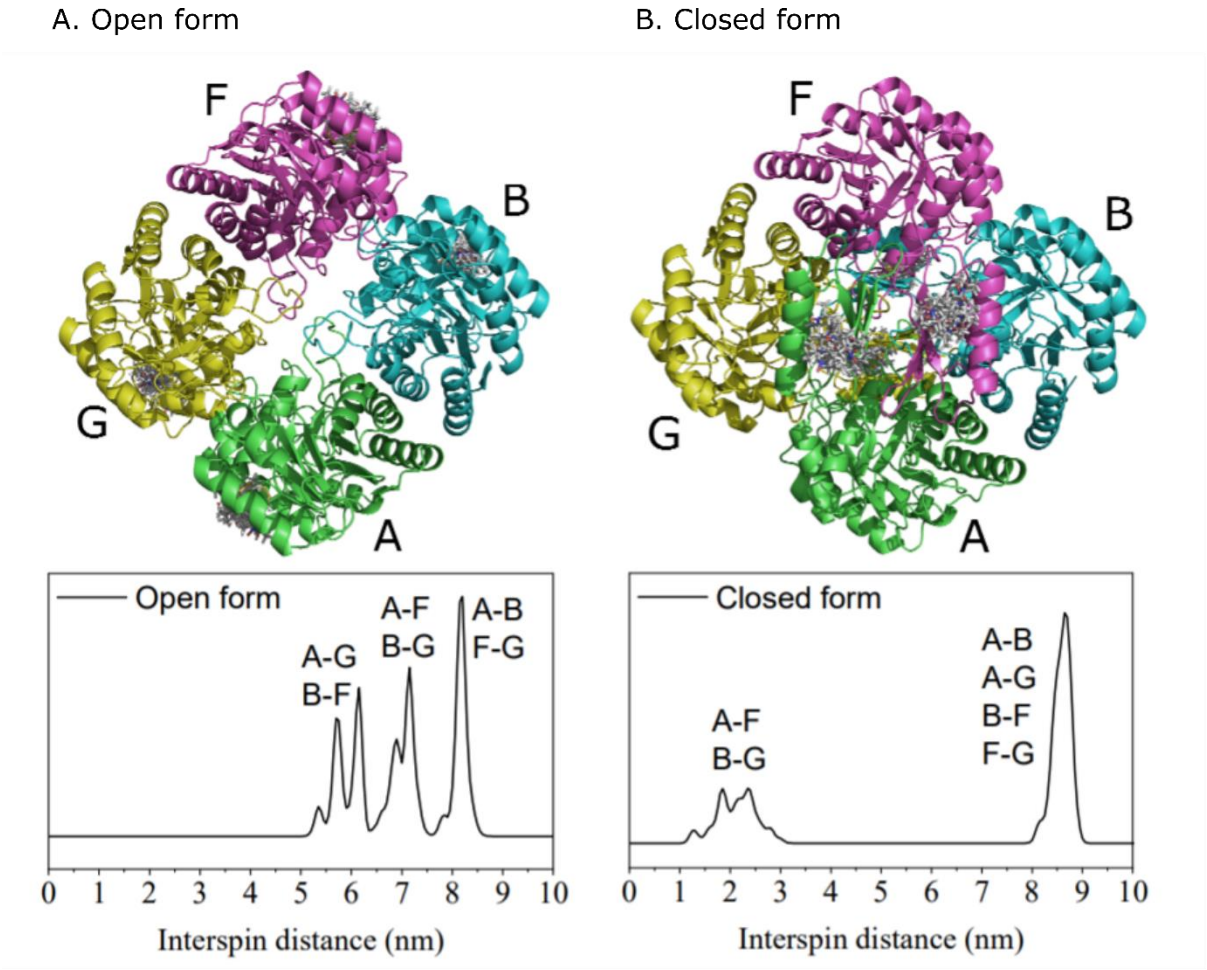


Figure 5.17. MMM analysis of MTSSL labelled *Tma*DAH7PS. Spin-label distance distribution predicted by using the possible rotamer library approach with the modelling software MMM. Rotamers of MTSSL attached on Cys28 are shown as sticks. Each protein subunit is coloured and labelled separately. A. Open-form prediction was based on coordinates from PDB 1RZM. B. Closed-form prediction was based on PDB 3PG9.

5.2.5. PELDOR measurements

The MTSSL-labelled *Tma*DAH7PS^{C58S} sample was buffer exchanged into D_2O buffer containing 50 mM BTP (pD 7.5), 100 mM KCl, and 200 µM PEP and sent to University of Manchester for PELDOR measurements. Sample was prepared by centrifugation of 70 µL of

protein mixed with 30 µl d6-glcerol. Then the sample was placed inside a 4 mm o.d. quartz EPR tube and heated at 60 °C for 5 minutes, followed by shock freezing in liquid nitrogen. The heat treatment step was due to consideration of the thermophilic nature of the protein. PELDOR measurements were performed at 50 K using a Bruker Elexsys 580 spectrometer for 24 hours. Software DEERAnalysis2013.2 was used for data processing and subtraction of background factor caused by intermolecular interactions.[252]

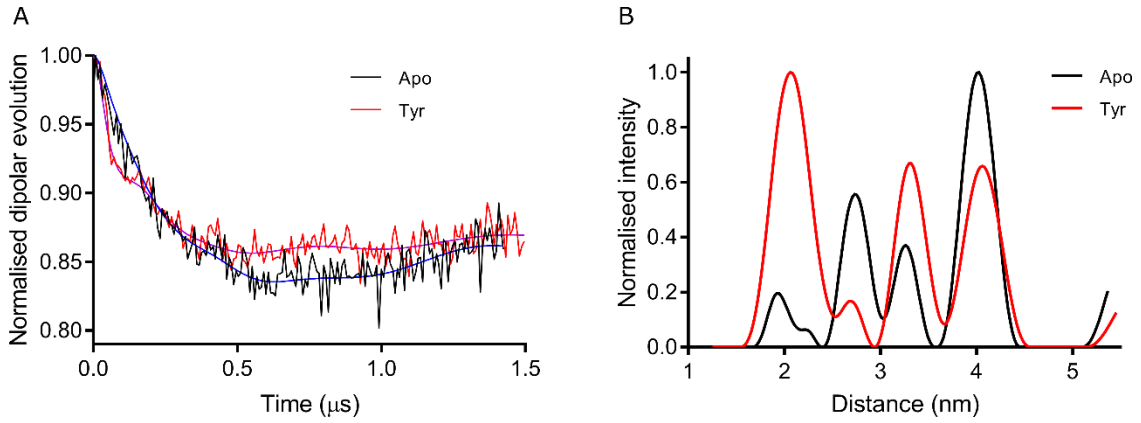


Figure 5.18. PELDOR data and distance distributions obtained from MTSSL-labelled *Tma*DAH7PS. A. Background corrected PELDOR traces shown as black (apo) or red (Tyr) line. Form factor fits are shown as blue (apo) or pink (Tyr) line. B. Distance distribution derived using Tikhonov regularisation shown as black (apo) or red (Tyr) line.

Data collected were of sufficient quality to derive reliable mean and width from the distribution based on the acquisition window. Distances beyond 5 nm were not accessible due to the limitation of the PELDOR spectrometer used in measuring long-range distances. Both PELDOR traces in the absence and presence of Tyr showed shallow oscillations, indicative of generally wide distributions of conformational states (Figure 5.18). In the absence of Tyr, the extracted distance distribution displayed a dominant mean distance at 4.03 nm. Two other smaller populations were also identified with mean distance of 3.26 and 2.73 nm. The least

dominant distance population peaked at 1.92 nm. These distance distributions were not present in the MMM prediction (Figure 5.17), suggesting that the protein naturally adopted various conformational states that were in between of the crystal extremes with the spin-labelled sites in the ~3-4 nm range of each other. Additionally, the presence of a small population at 1.92 nm, consistent with the MMM prediction for the opposite chains in the closed form, indicated that 'closed-like' conformations existed even in the absence of Tyr.

In the presence of 1 mM Tyr, although similar population maximums were observed at 4.06, 3.30 and 2.69 nm compared to the apo experiments (Figure 5.18), a significant shift in population towards the small distance at 2.06 nm was detected. This was consistent with the prediction from MMM analysis, in which the distance between spin-labelled sites from opposite chains ranged from approximately 1.5-2.5 nm (A-F and B-G, Figure 5.18). This observation verified that the presence of Tyr promoted formation of the inhibited closed form of the protein, resembling the crystal structure. However, the occurrence of distance populations at ~3-4 nm suggested that some proteins adopted 'closed-like' or 'open-like' conformations regardless the presence of absence of Tyr.

5.3. Summary and discussion

Studies conducted in this chapter explored the feasibility of employing FRET and EPR techniques in the multi-oligomeric *Tma*DAH7PS and *Gsp*DAH7PS model systems. Protein variants and tailored labelling strategies for both techniques were successfully established and executed. Although the mTG-mediated labelling for FRET posed multiple problems in labelling efficiency and protein stability of *Gsp*DAH7PS, the approach was proven to be a highly specific and sufficient labelling method, which could be further investigated in different

protein systems or altered labelling sites of *Gsp*DAH7PS. Future directions for specific labelling of *Gsp*DAH7PS may include identification of alternative labelling sites on the regulatory CM domains that leave the N-terminal flexible loop undisturbed and simultaneously satisfy appropriate FRET distances. For example, the α1 helix on the CM domain could be mutated to include a reactive Gln residue for mTG recognition, or a spacer sequence might be inserted at the N-terminus between the Q-tag and the beginning of first α helix. Extensive screening for reactive and undisruptive Gln-containing insertions might be involved.

On the other hand, albeit unspecific labelling of *Tma*DAH7PS caused high background FRET, a successful labelling scheme was established after various tests and alterations. The protein was effectively labelled with a pair of fluorophores with undisrupted functions. Single molecular FRET could be the next step towards measuring specific intramolecular FRET of the labelled *Tma*DAH7PS system in response to Tyr and differentiating that from non-specific intermolecular FRET. A more specific and orthogonal labelling approach without using enzymatic labelling could include the application of the fluorogenic addition reaction (FIARe) labelling strategy using a dimaleimide fluorophore.[253-254] This approach relies on use of fluorophores that bear two maleimide functional groups, which react with target protein containing appropriately spaced Cys residues through thiol addition reactions. The probes only fluoresce upon occurrence of the FIARe (Figure 5.19). Although it is a relatively new technique, the principles behind it have been demonstrated by successful testing on labelling of the epidermal growth factor receptors.[253] This approach can also be combined with mTG-mediated Q-tag labelling for other protein systems to achieve elegantly specific and orthogonal labelling.

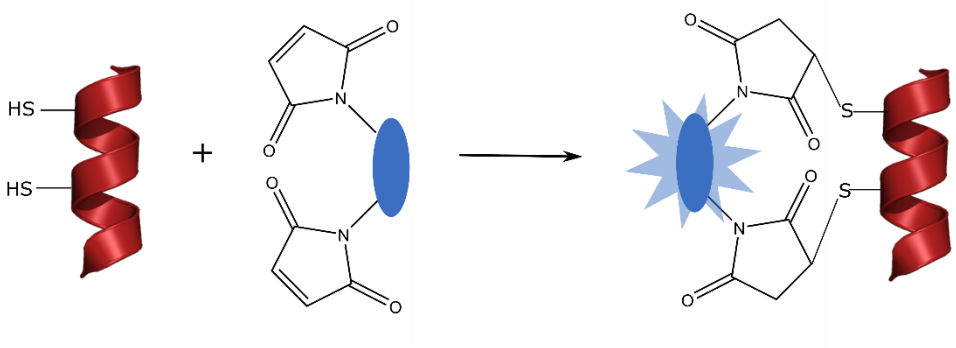


Figure 5.19. FlARe protein labelling strategy using a dimaleimide fluorophore.

The differences between PELDOR distributions of apo and Tyr-bound *Tma*DAH7PS show a clear trend in which the closed-form conformation becomes dominant upon addition of Tyr, and multiple conformational states that are different from the crystal structures constantly exist in the presence or absence of Tyr. This result is complementary to the molecular dynamic (MD) simulations performed by Dr Eric Lang.[255] The broadly open conformation obtained from the crystal structure did not maintain for longer than a few nanoseconds during the simulation (Figure 5.20). The ACT domains immediately deviated from the starting conformation and began exploring the nearby space. This information combined with the PELDOR results suggest that the wide-open conformation observed from the crystal structure is unstable in solution and may only exist under crystallisation constraints possibly due to the intermolecular interactions described in Chapter 2.3.3.

In contrast, the closed form remained stable over the course of MD simulations even in the absence of Tyr, with small displacements within 2 to 3 Å away from the original conformation, consistent with PELDOR distance distributions. The small displacements away from the closed-form crystal structure as observed from both MD simulations and PELDOR measurements support the incomplete inhibition of *Tma*DAH7PS activity by Tyr (Chapter 2),

in which a portion of the protein population is not fully closed allowing substrate access to the active site. MD simulations support that the 'closed-like' populations detected by PELDOR in the absence or presence of Tyr represent low-energy states that are likely to exist in solution. Future PELDOR measurements may consider conducting MMM calculations using coordinates from MD simulations for more accurate and realistic distance predictions.

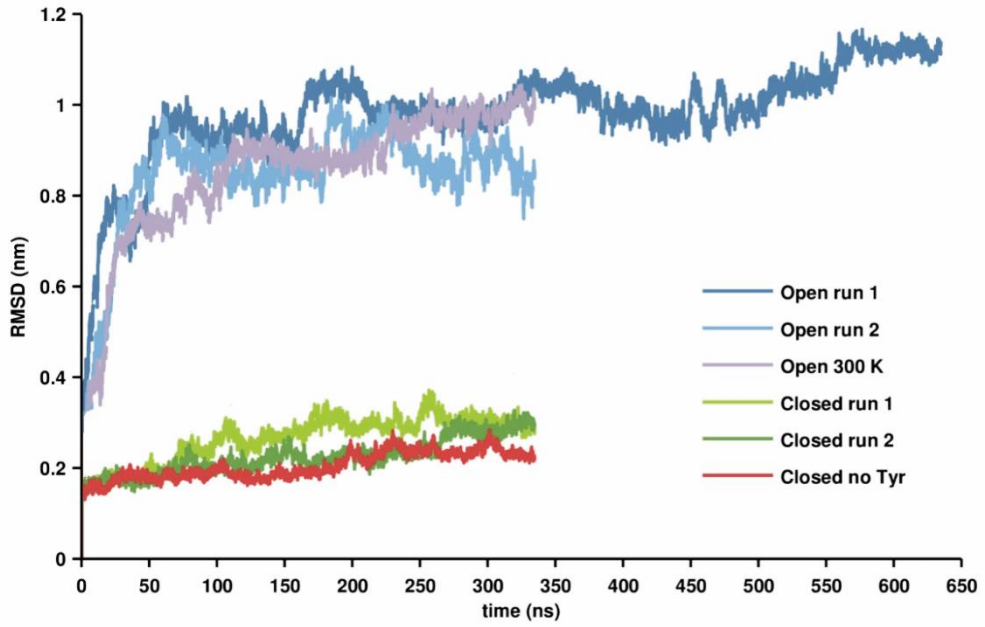


Figure 5.20. RMSD traces of C_α atoms in *Tma*DAH7PS from MD trajectories. Open form simulations were calculated based on crystal structure PDB 1RZM. Closed form simulations were calculated based on crystal structure PDB 3PG9. Figure was adapted from Lang.[255]

Chapter 6. Summary and remarks

6.1. Conformational equilibrium in Type Iβ DAH7PS

This thesis presents detailed functional and structural studies associated with allosteric regulation of several DAH7PS and CM enzymes and expands our current understanding of the remarkably diverse allosteric mechanisms employed by the DAH7PS family.[80] This diversity in allostery is largely related to the structural variations within the protein family.

The discrete ACT domain attached to the uninterrupted catalytic barrel of the Type Iβ DAH7PSs is responsible for the binding of allosteric effectors, which are the end-products of the shikimate pathway including Tyr and Phe (Chapter 2). In the absence of allosteric effectors, the protein adopts a relatively open and flexible active form with the ACT domains randomly exploring surrounding areas. In the presence of Tyr or Phe, the conformational population of the protein shifts towards more compact and rigid forms. ACT domains from diagonally opposite chains form interactions mediated by the allosteric ligand and physically hinder substrate access to the active site. Results from SAXS, PELDOR experiments and MD simulations (Chapter 5) on *Tma*DAH7PS and/or *Tye*DAH7PS variants, suggest that the apo protein can sample conformations that are 'closed-like' forms. Furthermore, in the presence of ligand, part of the population still adopts 'open-like' conformations rather than the fully closed inactive form observed in the crystal structure. The occurrence of 'open-like' conformations in the presence of ligand is also supported by the observed incomplete inhibition of catalysis. These results indicate that the allosteric mechanism of the ACT-DAH7PS enzymes is likely based on conformational selection in which the allosteric ligand alters the conformational equilibrium and stabilises the inactive conformations.

A similar conformational equilibrium likely exists in the CM-DAH7PS enzymes, such as *Gsp*DAH7PS, in which binding of chorismate or prephenate stabilises conformations resembling the closed-form crystal structure (Chapter 3). However, due to the dimeric nature of the CM domain, the relationship between the CM and the DAH7PS domains appears more intimate compared to the ACT-DAH7PS system, therefore the conformational freedom that the protein can explore may be less than those with the ACT domain. Overall, multiple lines of evidence have suggested that Type Iβ DAH7PSs with the discrete ACT or CM domains explore more diverse conformations in solution than those observed in the crystal structures. The presence of ligand is not required to initiate conformational changes in these DAH7PS enzymes, and sampling of intermediate states is reduced by the presence of allosteric ligand, which stabilises closed conformations.

6.2. Ser/Ile31 determines allosteric ligand selectivity in ACT-DAH7PS

Characterisation of *Tye*DAH7PS demonstrates that the ACT domain mediated allosteric regulation of DAH7PS, first discovered in *Tma*DAH7PS,[63] is not unique, and the conformational change upon ligand binding is likely a common allosteric mechanism employed by the ACT-DAH7PS enzymes. Investigations into the important interactions between the allosteric ligand and the ACT domain show that the presence of a key residue Ile at position 31, located at the ACT domain antiparallel β-sheets interface of *Tye*DAH7PS, results in a Phe-inhibited protein. In contrast, residue Ser31 in *Tma*DAH7PS stabilises the hydrogen bonding interactions with inhibitor Tyr. Studies on complementary mutations of I31S in *Tye*DAH7PS and S31I in *Tma*DAH7PS reinforce that the hydroxyl side-chains of Ser31 is

a key determinant for selective binding and potency of Tyr, whereas the hydrophobicity introduced by Ile is essential for binding of Phe at the equivalent site. These results combined with multiple sequence analysis of the ACT-DAH7PS family indicate that proteins with the 'His-X-Ser' motif at the ligand binding site of the ACT domain are responsible for reporting cellular level of Tyr, whereas proteins with the 'His-X-Ile' motif are often feedback regulated by Phe. Variations observed in SAXS profiles of *Tma*DAH7PS, *Tye*DAH7PS and their variants also support the multiple conformational states adopted by the ACT domain in solution due to its high degree of freedom. The conformation variations also indicate that features of the ACT domain, especially at the interfaces where ligand binding usually occurs, may affect the sampling of various energy states and likely plays a role in determining the conformational equilibrium between open and closed forms in the absence or presence of ligand, therefore determining the average population distribution.

6.3. Allosteric regulation can be interchanged

Exchanging the regulatory domains between *Tma*DAH7PS and *Gsp*DAH7PS results in two new functional proteins with interchanged allosteric behaviour, one of which is additionally a new bifunctional enzyme, and the other displays remarkably enhanced thermostability adopted from the more thermophilic element of the chimera (Chapter 3). Statistical coupling analyses on multiple protein systems provided the foundation and proposed the feasibility of engineering artificial allosteric systems. These analyses identified networks of coevolving amino acids that were involved in communication of allosteric signal in modular proteins and these networks showed strong connectivity.[144, 155] The full amino acid networks involved in the *Tma*DAH7PS and *Gsp*DAH7PS systems are yet to be defined, however, the interchangeable allostery and the

significant overlap between catalytically important residues and residues involved in allostery, identified in Chapter 3, shed light on the evolution of the residues associated with allostery.

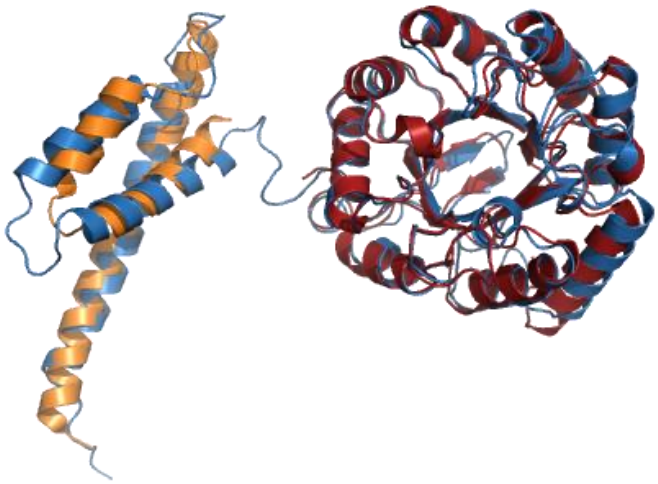


Figure 6.1. Superposition of *Pfu*CM (orange, PDB 1YBZ), and *Pfu*DAH7PS (red, PDB 1ZCO) with *Gsp*DAH7PS (blue).

Another protein chimera could be developed by fusion of *Pfu*CM as a putative regulatory domain on to the unregulated *Pfu*DAH7PS. Sequence and structural analyses of *Pfu*DAH7PS in relation to the requirements of allostery identified in Chapter 3, show that *Pfu*DAH7PS is equipped with allosterically important residues. Additionally, the tetrameric structure of *Pfu*DAH7PS provides a platform ready for adoption of allosteric regulation.[117] On the other hand, although *Pfu*CM shares high structural similarity with the CM domain of *Gsp*DAH7PS (Figure 6.1), the α2-α3 loop and the third helix (α3) of *Pfu*CM appear slightly shorter compared to those of *Gsp*CM. These regions are important in interactions with the DAH7PS barrels. Careful design of the loop, and the linker region connecting the *Pfu*CM and *Pfu*DAH7PS is likely to be crucial for successful construction of the chimera. The ^{Pfu}CM-^{Pfu}DAH7PS fusion may provide insights into the evolutionary process in which gene fusion events between the regulatory and the catalytic components led to acquisition of allostery in contemporary Type

Iβ DAH7PS enzymes, and the modifications required for addition of regulatory properties on the discrete CM enzyme.

A trade-off between catalytic activity and allostery is evident in the Type Iβ family, and the degree of attenuated activity depends on the nature of the regulatory domains. Removal of the regulatory domains from both *Tma*DAH7PS and *Gsp*DAH7PS resulted in catalytically enhanced proteins.[63, 78] The ^{Gsp}CM-^{Tma}DAH7PS chimera with the more restricted dimeric CM domain displayed impaired catalytic activity compared to the wild-type *Tma*DAH7PS with the mobile ACT domains. In contrast, the ^{Tma}ACT-^{Gsp}DAH7PS chimera displayed a two-fold boost in catalysis compared to the wild-type *Gsp*DAH7PS. This trade-off may be linked to allosteric regulation utilising conserved catalytic residues as detailed in Chapter 3.

6.4. Type II DAH7PS and AroQ CM

Although it was proposed that the formation of protein complex between AroQ CM and Type II DAH7PS as observed in *M. tuberculosis* might be a common allosteric strategy utilised by all AroQ CM and Type II DAH7PS enzymes,[30] this hypothesis could not be validated with the *Hpy*DAH7PS system (Chapter 4). *Hpy*DAH7PS displays different allosteric regulation compared to *Mtu*DAH7PS. It exhibits only a moderate level of inhibition by aromatic amino acids at low substrate concentrations, and adopts different oligomeric structures and quaternary arrangement compared to *Mtu*DAH7PS. Overall, the allosteric regulation employed by Type II DAH7PS appears to vary in different organisms, however, the structural extensions at the dimer and tetramer interfaces of these proteins likely play an important role in allosteric ligand binding. The allosteric mechanisms of Type II DAH7PSs do not display dramatic conformational changes compared the Type Iβ, but are conferred via subtle changes in dynamic

communications. The AroQ CM from *H. pylori* also displays substantial functional and structural differences compared to the *Mtu*CM. It is the only CM in the *H. pylori* genome and has normal-range catalytic activities, therefore does not require activation by *Hpy*DAH7PS. Successful crystallisation of *Hpy*CM reveals the elongated C-terminal helix that does not fit into the groove on top of the *Hpy*DAH7PS barrels and severe clashes could occur at the interfaces between the two proteins. In contrast, the equivalent helix in *Mtu*CM plays an important role in formation of the protein complex between *Mtu*CM and *Mtu*DAH7PS.[30]

6.5. The role of oligomerisation in allostery and protein stabilisation

Variations in quaternary structure of the DAH7PS family are evident. Whilst many members of the Type Iβ DAH7PS are tetrameric (Chapter 2 and 3), *Tye*DAH7PS displays an oligomeric equilibrium between dimer and tetramer. Besides altering the equilibrium between active and inactive conformations of *Tye*DAH7PS, the presence of allosteric inhibitor Phe also appears to affect the oligomeric equilibrium by shifting the position of the equilibrium towards tetramer formation. The *Hpy*DAH7PS from Type II family also displays dimer-tetramer equilibrium (Chapter 4), and has a weaker tetramer interface compared to the Type II DAH7PS from *M. tuberculosis*, which is predominantly tetrameric with minor dimeric species only observed at very low concentrations.[207] It is not yet known whether these dimeric species observed in both Type Iβ and Type II families exist in the cell, or if they have any physiological relevance. However, the relatively weak tetramer interfaces and the interface modifications identified in both *Tye*DAH7PS and *Hpy*DAH7PS may provide insights into the evolution of their quaternary structure. Oligomerisation appears to be a prerequisite for allostery of DAH7PS that involves dimerisation of the regulatory module. For instance, removal of regulatory ACT domain from

*Tma*DAH7PS and CM domain from *Gsp*DAH7PS both produced functional dimers with no allosteric behaviour.[63, 78] Additionally, single amino acid substitutions at tetramer interface of *Mtu*DAH7PS were able to render the protein dimeric, and the dimers were functional with significantly impaired allosteric regulation.[140] Therefore, it appears that the formation of tetrameric species is to provide allostery in regulated Type Iβ and Type II DAH7PSs. If the observed dimeric species exist in the cell, it may be possible that the protein is en route to modify the level of existing allosteric regulation by evolving weaker tetramer interfaces to produce dimers that are not subject to immediate allosteric regulation. This hypothesis may be linked to the moderate level of inhibition displayed by *Hpy*DAH7PS (Chapter 4) and the increase in tetramer formation of *Tye*DAH7PS in the presence of Phe (Chapter 2).

In addition to the delivery of allosteric regulation, oligomerisation in some members of the DAH7PS family also serves a role in stabilisation of the protein. It has been reported that an unregulated member of Type Iβ DAH7PS from *P. furiosus* was tetrameric, and rendered dimeric by a single amino acid substitution at the tetramer interface. The dimers were catalytically more active, however their thermostability was significantly compromised compared to wild-type protein.[134] In another example, the tetrameric interface of Type Iα DAH7PS from *N. meningitidis* was disrupted and resulted in dimers that were catalytically active and allosterically inhibited to the same extend as the wild-type protein, however, showed reduced thermal stability.[139] It seems that whether oligomerisation primarily facilitates allostery and/or provides protein stability, depends on the nature of the allosteric system. If the delivery of allostery requires the discrete regulatory domains (in the case of regulated Type Iβ DAH7PS), or binding of allosteric ligands occurs at the tetramer interfaces (in the case of Type II DAH7PS), the tetramer formation is unequivocally crucial for allostery. However, it is not ruled out that oligomerisation may also have a role in protein stabilisation in such cases. For DAH7PS enzymes that display no regulation (in the case of unregulated Type Iβ DAH7PS) or

regulation that is not directly dependent on dimerisation of the regulatory elements (in the case of Type Iα DAH7PS), tetramer formation appears to have the primary role of protein stabilisation.

6.6. Evolution of DAH7PS

Based on previously reported studies and the results presented in this thesis,[62-63, 72, 78, 117, 134] a fuller model for the evolution of allostery in DAH7PS can be proposed (Figure 6.2). It has been suggested that the ancestral DAH7PS may be a dimer incapable of allosteric regulation.[72] Association of the dimers into tetramers, such as *Pfu*DAH7PS, was likely followed by the acquisition of allosteric regulation via gene fusion events with the ACT or CM domains, leading to the formation of the Type Iβ DAH7PS family (Figure 6.2). The Type Iα and Type II DAH7PS families share similar catalytic barrel structures, but low sequence identity with the Type Iβ. They are also equipped with N-terminal structural extensions that may lead to tetramer formation. Interestingly, the Trp binding site at the tetramer interface of Type II DAH7PS (such as *Mtu*DAH7PS) displays vague topological similarity to a CM domain, which has strong predilection for dimerisation. Additionally, the N-terminal extension and insertion of the Type Iα DAH7PS (such as *Nme*DAH7PS and *Eco*DAH7PS) display features resembling the ACT domain and result in weaker tetramer interfaces compared to the Type II. It is yet to be determined whether the excision of the N-terminal extension from its host would still fold into a compact structure, maybe given its length, into an ACT-like or CM-like structure. However, it is possible that the Type Iα and Type II were evolved from Type Iβ, with truncated, differently folded or smeared out extensions that evolved from the ACT or CM domains respectively. More sophisticated allosteric communication networks were subsequently evolved in Type Iα and Type II families than the gating mechanism of the Type Iβ family.

Further phylogenetic studies are necessary to fully understand the evolutionary relationships among the DAH7PS families.

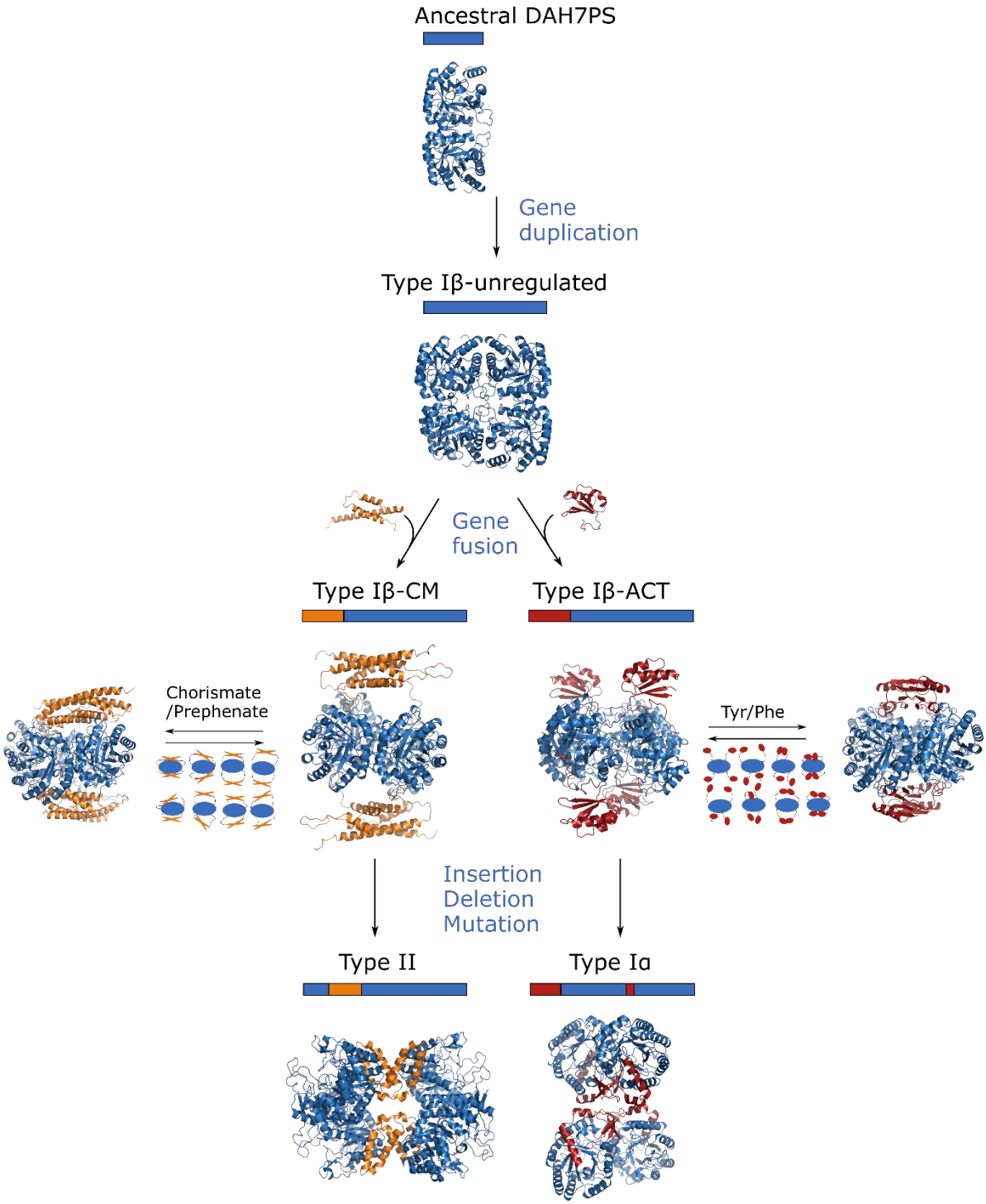


Figure 6.2. Schematic representation of proposed evolutionary relationships among DAH7PS subfamilies. Conformational equilibrium in the regulated Type Iβ family is also shown. The DAH7PS

barrel is shown in blue. The CM and the ACT domain in Type Iβ DAH7PS are shown in orange and red, respectively. The insertions in the DAH7PS barrel of Type II are shown in orange, the insertions and extensions in the DAH7PS barrel of Type Iα are shown in red.

6.7. Conclusion

The DAH7PS enzyme family, situated at a pivotal gateway, links primary metabolism of carbohydrates with the aromatic amino acids pathways en route to biosynthesis of numerous metabolites. Presented with such an important role, it is truly fascinating that the DAH7PS family has adapted remarkably disparate allosteric networks in response to cellular demand in different organisms under various selection pressure. Believed to have evolved from an unregulated enzyme, the Type Iβ DAH7PSs fused with ACT or CM terminal domains employ simple gating mechanisms associated with large conformational changes to control catalysis. Many conserved catalytic residues appear to have been repurposed with additional allosteric roles. Studies on several protein systems included in this thesis reveal the truly dynamic conformational equilibrium that lies behind the conformational changes, which extends far beyond the crystallographic snapshots. This study also demonstrates that members of DAH7PS from the same subgroup can adopt variations in conformations and allostery, albeit sharing similarities in many aspects such as their ligand binding motif in the ACT domain. Regulation by the intermediate product of the shikimate pathway, is interchangeable with, feedback regulation by pathway end-products, demonstrating the ease and portability of which different allosteric strategies can be acquired by the organism. This finding also contributes to the understanding of the amino acid networks associated with allostery. The discrete regulatory domains were possibly modified and repurposed in different organisms during evolution, leading to the establishment of Type Iα and Type II DAH7PS families, in which subtle and

sophisticated allosteric networks without dramatic conformational changes have been developed. The methodologies and techniques developed in this study for the examination of the dynamic allostery in DAH7PS model systems can provide valuable resources for future investigations of allostery in other proteins.

Chapter 7. Experimental procedures

7.1. General methods

7.1.1. Sequence analysis

Sequence alignments figures were generated with Clustal Ω [256] and formatted with ESPript 3.[257] For Chapter 3, total of 2523 sequences from the DAH7PS family of interest (Type Iβ) were obtained from Pfam (PF00793)[258], aligned using SeaView 4[259] and conserved residues were identified and graphed with WebLogo3.[120]

7.1.2. Protein structure images

Structural representation of proteins were generated using the PyMOL Molecular Graphics System (version 1.8.2.1, Schrödinger, LLC).[260]

7.1.3. Water

Water used in all experiments was purified by a Millipore Milli-Q® system. For molecular biology and cell culture related experiments, water was sterilised by autoclaving prior to use. For EPR measurements, heavy water (D_2O) was used.

7.1.4. Buffer solutions

All standard buffers and solutions used were made with Milli-Q water, and the pH values were determined using a Mettler Toledo™ S220 SevenCompact™ pH/Ion meter and adjusted by using HCl or NaOH solutions. The pD values for buffers made with heavy water were adjusted with the equation pD = pH + 0.4, where pH was the observed value.

When metal ions were required to be removed from buffer solutions, such as for metal selectivity measurements, Chelex® 100 resin (Bio-Rad) was incubated with the target solution overnight with gentle stirring. The mixture was filtered using a 0.22 μm membrane prior to use.

7.1.5. Media

All standard *E. coli* cell cultures were grown in lysogeny broth (LB) at 20 $g \cdot L^{-1}$ (unless otherwise stated) made up with Milli-Q water and sterilised by autoclaving. Appropriate antibiotics were added before use. For agar plates, LB-agar media (Miller's) was dissolved in water at 37 $g \cdot L^{-1}$ before sterilisation by autoclaving. The appropriate antibiotics were added after the media was cooled to desired temperatures, and the mixture was subsequently plated in petri dishes. SOC media used during transformation consisted of 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 20 mM glucose.

7.2. Cloning

7.2.1. Genes

All genes encoding wild-type proteins of interest were synthetic genes purchased from GeneArt and codon optimised for expression in *E. coli* cells. Protein variants were created from the wild-type genes by PCR techniques.

Plasmids containing genes of interest were isolated from overnight cell cultures by using the High Pure Plasmid Isolation Kit (Roche). Concentrations of DNA products were determined by measuring absorbance at 260 nm using NanoDrop® ND-1000 spectrophotometer.

Table 7.1. Primers used for cloning.

Protein	Primer 5'-3'
Chapter 2	
Cloning of *Tye*DAH7PS constructs	
*Tye*DAH7PS-His_6 (C-terminus, uncleavable)	FWD AGGAGATATACCATGGTGATTGTGATG REV GTGGTGGTGCTCGAGAATTTTTTCGGCAATGC
*Tye*DAH7PS (untagged)	FWD AGGAGATATACCATGGTGATTGTGATG REV TTATTAGTGGTGGTGCTCGAGAATTTTTTCGGCAATGC
His_6-*Tye*DAH7PS (N-terminus, cleavable)	FWD CGCGCGGCAGCCATATGGAAAACCTGTATTTTCAGGGC REV GGTGGTGGTGCTCGAGTCATCAAATTTTTTCGG
Mutagenesis	
*Tma*DAH7PS^{S31I}	FWD CTTGAAGTGTCACATTATTAAAGGTCAGGAAAG REV CTTTCCTGACCTTTAATAATGTGACACTTCAAG
*Tye*DAH7PS^{I31S}	FWD CTATAAACCGCATGTTTCTTATGGCACCACCCGT REV CGGGTGGTGCCATAAGAAACATGCGGTTTATAG
Chapter3	
Cloning of ^{Tma}ACT-^{Gsp}DAH7PS	
PCR 1-^{Tma}ACT domain and linker with overlap	FWD GCTTCGAAGGAGATAGAACCATGATTGTCGTTTTGAAACCC REV GACGAAATATTGATTGCCGTTTCCTATCTTCACATCT
PCR 2-^{Gsp}DAH7PS domain with overlap	FWD GTGAAGATAGGAAACGGCAATCAATATTTCGTCATGGG REV GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATGCACGAACAAACTGACG
PCR 3 -fusing the above products	FWD GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG REV GGGGACCACTTTGTACAAGAAAG
Cloning of ^{Gsp}CM-^{Tma}DAH7PS	
PCR 1-^{Gsp}CM domain and linker with overlap	FWD GCTTCGAAGGAGATAGAACCATGGGTAATGAACGTCTGGAT REV CGCTATGATGGTGAAATACTGATTGCCATCACCAAT
PCR 2-^{Tma}DAH7PS domain with overlap	FWD AATCAGTATTTCACCATCATAGCGGGACCG REV GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAATTCACCTTCACCCCCAG
PCR 3-fusing the above products	FWD GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG REV GGGGACCACTTTGTACAAGAAAG
Chapter 4	
Cloning of *Hpy*DAH7PS	
His_6-*Hpy*DAH7PS	FWD GGCAGCGGCGCGATGTCAAACACAAC REV GAAAGCTGGGTGTCAAGTGCGTTGTT
Chapter5	
Cloning of Q-tagged *Gsp*DAH7PS	
PCR 1-Q-tag	FWD ATGTGGGCGCTGCAGCGTCCGCATGGTAATGAACGTCTGGAT REV TTATTACTCGAGTGCACGAACAAACTGACG
PCR 2-In-Fusion®	FWD CGCGCGGCAGCCATATGGAAAACCTGTATTTTCAGGGCAGCGGCGCGATGTGGGCGCTGC REV GGTGGTGGTGCTCGAGTTATTACTCGAGTGCACGAACAAAC
Mutagenesis	
*Gsp*DAH7PD^{D302C}	FWD GCACCGGTCGTCGTTGCCTGCTGATTCCGTGTG REV CACACGGAATCAGCAGGCAACGACGACCGGTGC
*Gsp*DAH7PD^{C307S}	FWD GATCTGCTGATTCCGAGCGCAAAAGCAGCACTG REV CAGTGCTGCTTTTGCGCTCGGAATCAGCAGATC
*Gsp*DAH7PD$^{D302C-C307S}$ (*Gsp*DAH7PD^{D302C} as template)	FWD TGCCTGCTGATTCCGAGCGCAAAAGCAGCACTG REV CAGTGCTGCTTTTGCGCTCGGAATCAGCAGGCA
*Tma*DAH7PS^{C58S}	FWD GTTCGAGTCGCTGGATAGCGTGGAAAGCGTTGTG

	REV CACAACGCTTTCCACGCTATCCAGCGACTCGAAC
*Tma*DAH7PS$^{C58S-C235S}$ (*Tma*DAH7PS^{C58S} as template)	FWD CACGAAGATCATACTGAGCGAAAGGGGAATCAGA
	REV TCTGATTCCCCTTTCGCTCAGTATGATCTTCGTG

7.2.2. Polymerase chain reaction

Primers used for polymerase chain reaction (PCR) were purchased from Invitrogen and diluted to desired concentrations using sterile water or TE buffer (10 mM Tris-Cl, 0.1 mM EDTA, pH 8.0). PCR experiments were performed using a Veriti® 96-well Thermal Cycler (Applied Biosystems). Phusion® High-Fidelity DNA Polymerase was generally used unless otherwise stated.

Colony PCR experiments were usually performed to verify the size of the gene product. Small portions of cell colonies were collected from agar plates and used as templates for PCR. A pair of primers including a gene specific primer and a plasmid specific primer, and Taq DNA polymerase were used. An optional elongated denaturation step was sometimes added at the start of the reaction to allow sufficient cell lysis.

7.2.3. Gateway® cloning

Gateway recombination technology is a universal clothing method that allows fast and efficient transfer of DNA segments into various vectors in fewer steps compared to traditional restriction enzyme cloning.[261] Protein constructs including ^{Tma}ACT-^{Gsp}DAH7PS, ^{Gsp}CM-^{Tma}DAH7PS, and *Hpy*DAH7PS were cloned using this technique (Table 7.1). Specific recombination sites were added to the amplified target gene using PCR. Optional TEV cleavage site was included when required. Linear DNA product was transformed into pDONR^{TM}221 vector via BP clonase reaction and into TOP10 competent cells following the product manual. The pDONR^{TM}221 vector permits recombination cloning into multiple destination vectors allowing optimal

expression of the gene of interest. The resulting clone was plated on LB-agar plates with Kanamycin, from which cell colonies were selected and used for plasmid extraction. The purified plasmid was used for the LR clonase reaction and transformed into pDEST^{TM}14, pDEST^{TM}15, or pDEST^{TM}17 depending on requirement for a tagless, GST tag, or His_6 tag, respectively (Table 7.2). Subsequent transformation procedures into expression vectors followed standard method described below.

Table 7.2. Proteins used in this study.

Protein	Vector	Expression cell line (*E. coli*)
*Tye*DAH7PS-His_6	pET28-b(+)	BL21*(DE3)
*Tye*DAH7PS (untagged)	pET28-b(+)	BL21*(DE3)
His_6-*Tye*DAH7PS	pET28-a(+)	BL21*(DE3)
*Tye*DAH7PS^{I31S}	pET28-b(+)	BL21*(DE3)
*Tma*DAH7PS	pT7-7	BL21*(DE3)
*Tma*DAH7PS^{S31I}	pT7-7	BL21*(DE3)
*Tma*DAH7PS^{S31V}	pT7-7	BL21*(DE3)
^{Tma}ACT-^{Gsp}DAH7PS	pDEST14	BL21*(DE3)
^{Gsp}CM-^{Tma}DAH7PS	pDEST15	BL21(DE3)pLysS
*Gsp*DAH7PS	pET28-b(+)	BL21*(DE3)
*Hpy*DAH7PS	pDEST17	BL21*(DE3)
*Hpy*CM	pDEST15	BL21*(DE3)
*Pfu*CM	pDEST15	BL21*(DE3)
*Gsp*DAH7PS^{D302C}	pET28-b(+)	BL21*(DE3)
*Gsp*DAH7PS^{C307S}	pET28-b(+)	BL21*(DE3)
*Gsp*DAH7PS$^{D302C-C307S}$	pET28-b(+)	BL21*(DE3)
*Gsp*DAH7PS$^{D302C-C307S-Q}$	pET28-b(+)	BL21*(DE3)
*Tma*DAH7PS^{C58S}	pT7-7	BL21*(DE3)
*Tma*DAH7PS$^{C58S-C235S}$	pT7-7	BL21*(DE3)
mTG	pDJ1-3	BL21*(DE3)

7.2.4. In-Fusion® cloning

In-Fusion® HD cloning (Clontech) is a rapid cloning method allowing directional cloning of DNA fragments into various vectors by enzymatic recognition of a 15-basepair overlap engineered at the ends of the target DNA. Constructs for *Tye*DAH7PS were cloned using this technique (Table 7.1, 7.2). Target vector was linearised using restriction enzyme digest and purified by NucleoSpin® gel purification kit. Restriction sites NcoI and XhoI were used for

the construct *Tye*DAH7PS-His_6 (C-terminus), and restriction sites NdeI and XhoI were used for His_6-*Tye*DAH7PS (N-terminus). Gene-specific primers were designed for PCR to incorporate the recognition sites at both ends of the gene based on selected restriction sites (Table 7.1). The PCR product, linearised vector, and In-Fusion® enzyme mixture were incubated and transformed as per manufacturer's instructions.

7.2.5. Site-directed mutagenesis

QuikChange® site-directed mutagenesis technique was used to generate protein variants including *Tma*DAH7PS^{S31I}, *Tma*DAH7PS^{S31V}, *Tma*DAH7PS^{C58S}, *Tma*DAH7PS$^{C58S-C235S}$ *Tye*DAH7PS^{I31S}, *Gsp*DAH7PS^{D302C}, *Gsp*DAH7PS^{C307S}, *Gsp*DAH7PS$^{D302C-C307S}$. Primers for mutagenesis were designed with the point of mutation in the centre and sufficient overlaps (~15 base pairs) on both ends (Table 7.1). Mutagenesis was performed using a QuikChange II Site-Directed Mutagenesis Kit (Agilent), using cycling protocols and reaction components (50 µL) recommended by the manufacturer. PCR products were treated with Dpn1 (New England BioLabs) for removal of methylated templates.

7.2.6. Agarose gel electrophoresis

Agarose gels were prepared by mixing 1% (w/v) LE agarose (Seakem) with Tris-acetate EDAT (TAE) buffer (50 mM Tris, 20 mM acetic acid, and 1 mM EDTA) and heating until all agarose was dissolved. SYBR® Safe DNA gel stain (Invitrogen) was then added to the mixture before the gel was casted. DNA samples were mixed with 6x loading buffer [60 mM Tris-HCl, 60 mM EDTA, 0.2% (w/v) Orange G, 0.05% (w/v) xylene cyanol FF, 60% (v/v) glycerol] and loaded in each well. The electrophoresis was conducted at 85 V in TAE buffer for 45 minutes or until the dye front reached the desired position on the gel.

7.2.7. Transformation

Chemically competent cells TOP10 (Invitrogen) and Stellar^{TM} (Clontech) were used for plasmid propagation, and *E. coli* BL21* (DE3) or BL21* (DE3) pLysS cells were used for protein expression. An aliquot of the competent cell (50 µL) was thawed on ice and gently mixed with the target plasmid (~100-200 ng) before incubation on ice for 30 minutes. A heat shock of 42 °C for 45-60 seconds was used, followed by further incubation on ice for 2 minutes. SOC media (250 µL) was added into the cell mixture and this was incubated at 37 °C with shaking at 200 rpm for one hour. The resulting cell was plated on LB-ager plates with appropriate antibiotics and incubated overnight at 37 °C to produce colonies used for colony PCR, subsequent cell culture or plasmid extraction.

7.2.8. DNA sequencing

DNA sequencing service was provided by the Canterbury Sequencing Facility at the University of Canterbury, or Macrogen DNA Sequencing Service. Approximately 100 ng·µl^{-1} DNA product was supplied for each sample. Sequencing primers used for this study can be found in Table 7.3.

Table 7.3. Sequencing primers.

Sequencing primers 5'-3'	
M13	FWD GTAAAACGACGGCCAG REV AACAGCTATGACCATG
T7	FWD TAATACGACTCACTATAGGG REV CTAGTTATTGCTCAGCGGTG

7.3. Cell cultures

7.3.1. Protein expression

Precultures (50-200 mL culture in 200-500 mL flasks) were prepared by inoculation of LB media containing appropriate antibiotics with the expression cell strains, followed by overnight growth at 37 °C with shaking at 200 rpm. The precultures were then used to inoculate 1 L LB media in 3-L baffled flasks. Cell cultures were grown at 37 °C with shaking (200 rpm) until mid-logarithmic phase (OD_{600} 0.4-0.8 AU).

Most genes (except for mTG) used in this study were inserted in pDEST, pET or pT7-7 series expression vectors (Table 7.2). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used to induce protein expression with final concentration of 0.5 mM. The concentration of IPTG might be varied from 0.1 to 1 mM during test expressions. Following induction by IPTG, cell culture was incubated at 23 °C overnight or 37 °C for four hours with shaking at 200 rpm.

7.3.2. Autoinduction

Autoinduction of protein expression was used for mTG. The mTG proenzyme with C-terminal His_6 tag in the pDJ1-3 plasmid was transformed into *E. coli* BL21*(DE3). The preculture was grown in the ZYP-0.8G media under standard conditions as described in Chapter 5.1. ZYP-0.8G is a rich growth medium for production of high cell densities with little to no induction of expression.[241] A preculture (5 mL) was used to inoculate a large culture (500 mL) in ZYP-505 media. ZYP-505 is another rich growth medium with little to no induction of expression during log phase and allows autoinduction as the culture approaches saturation.[241] The inoculated culture was harvested after two hours incubation at 37 °C and overnight incubation at 22 °C with shaking at 220 rpm. The components of the autoinduction media are listed below.

ZY: 10 g N-Z-amine AS (or tryptone), 5 g yeast extract, 925 ml water

NPS: 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 50 mM Na_2HPO_4 (20 x NPS stocks were usually prepared)

505: 0.5% glycerol, 0.05% glucose, 0.2% α-lactose (100 x 505 stocks were usually prepared)

Metals mixture: 50 µM $FeCl_3$, 20 µM $CaCl_2$, 10 µM $MnCl_2$, 10 µM $ZnSO_4$, 2 µM $CoCl_2$, 2 µM $CuCl_2$, 2 µM $NiCl_2$, 2 µM Na_2MoO_4, 2 µM Na_2SeO_3, 2 µM H_3BO_3 (1000 x metals mixture stocks were usually prepared)

All metal stocks were prepared individually and autoclaved. $FeCl_3$ was dissolved in 50 mM HCl and filtered using a 0.22 µm membrane.

ZYP-0.8G recipe contained final concentrations of 1mM $MgSO_4$, 0.8% glucose, 1x NPS, 100 $\mu g \cdot mL^{-1}$ kanamycin, 25 $\mu g \cdot mL^{-1}$ chloramphenicol, and 50 $\mu g \cdot mL^{-1}$ ampicillin dissolved in ZY stock. $MgSO_4$ was added before NPS to avoid precipitation.

ZYP-505 recipe contained 1mM $MgSO_4$, 1 x metals mixture, 1x 505, 1x NPS, 100 $\mu g \cdot mL^{-1}$ kanamycin, 25 $\mu g \cdot mL^{-1}$ chloramphenicol, and 50 $\mu g \cdot mL^{-1}$ ampicillin all dissolved in ZY stock. $MgSO_4$ and metal mixture was added before NPS to avoid precipitation.

7.3.3. Cell harvesting

Large cell cultures (> 500 mL) were harvested by centrifugation in a Fiberlite™ F9-6x1000 LEX fixed-angle rotor (Thermo Fisher Scientific) using 1 L bottles at 14000 g for 30 minutes at 4 °C. Cultures less than 50 mL were harvested in sterile 50 mL tubes using a Fiberlite™ F14-14x50cy fixed-angle rotor (Thermo Fisher Scientific). Small cultures (< 5 mL) were

harvested in 1.5 mL micro-centrifuge tubes at 17000 g for 1 minutes. Harvested cell pellets were either immediately lysed or snap frozen and stored at -80 °C.

7.4. Protein purification

7.4.1. Cell lysis

Most cell lysis was performed using sonication, although chemical lysis and cell disruptor were also occasionally used. Cell pellets were resuspended in 30-40 mL chilled lysis buffer per litre of cell growth and sonicated on ice using an Omni-Ruptor 4000 Ultrasonic Homogeniser with 4–6 repeats of 5 minutes at 80% power with 40% pulsation. The lysis buffer contained 0.2 M Tris-Cl (pH 6.0) for mTG, or 50 mM BTP (pH 7.5), 1 mM EDTA, 100 mM KCl, 200 µM PEP, and 200 µM TCEP for other proteins.

Cell cultures with small volumes, such as test expression cultures, were chemically lysed using BugBuster® Protein Extraction Reagent (Novagen). Cell pellets were resuspended in lysis buffer containing 5% (v/v) BugBuster® and incubated at room temperature with gentle shaking for at least 20 minutes.

A cell disruptor (M-110P Microfluidiser®) was used to recover maximum level of soluble proteins from a large cell culture or when the yield obtained from sonication was not sufficient. Resuspended cell pellets were applied on the cell disruptor with chilled buffers at 18,000 psi for 3-5 times until the cell lysate reached ideal clarity and viscosity. All cell lysate was treated with Benzonase® for 10 minutes at room temperature before centrifugation at 40000 g for 30 minutes at 4 °C to remove the insoluble cell debris.

7.4.2. Heat treatment

Following cell lysis, heat treatment was performed as the first step of purification for proteins that tolerate high temperatures. These proteins include variants of *Tma*DAH7PS, *Gsp*DAH7PS and *Tye*DAH7PS. The heat treatment was conducted at 70 °C for 30 minutes for *Tma*DAH7PS variants and 60 °C for *Gsp*DAH7PS and *Tye*DAH7PS variants. The sample was then left to cool to room temperature and clarified by centrifugation.

7.4.3. Hydrophobic interaction chromatography

Proteins without purification tags, including *Tma*DAH7PS variants, ^{Tma}ACT-^{Gsp}DAH7PS, and the tagless *Tye*DAH7PS, were purified using hydrophobic interaction chromatography (HIC). 1 M $(NH_4)_2SO_4$ was slowly added into the crude supernatant following heat treatment, with gentle stirring at room temperature until dissolved. This solution was then filtered using a 0.22 µm membrane and loaded on a SOURCE™15Phe column (GE Healthcare) preequilibrated with 20 mM BTP (pH 7.5), 1 M $(NH_4)_2SO_4$. A linear gradient of 1 to 0 M $(NH_4)_2SO_4$ was applied over 70 mL elution. The fractions containing protein of interest were immediately collected and exchanged into low salt buffers to avoid protein precipitation.

7.4.4. Immobilised metal affinity chromatography

Proteins with engineered His_6 purification tag, including *Tye*DAH7PS variants, *Gsp*DAH7PS variants, *Hpy*DAH7PS variants, and mTG, were purified using immobilised metal affinity chromatography (IMAC). The 5 mL HiTrap™ HP (Ni^{2+}) columns (GE Healthcare) were used for IMAC purifications. The columns were equilibrated with binding buffer containing 20 mM BTP (pH 7.5), 20 mM imidazole, 150 mM NaCl (or 100 mM KCl), 200 µM PEP, and 200 µM TCEP (unless otherwise specified). Filtered protein samples were loaded on the column and

washed with at least 5 column volumes of binding buffer to remove unspecific binding of impurities. A gradient of 20 to 500 mM imidazole was applied over 50 mL for protein elution.

For mTG, activation of the enzyme by trypsin (Chapter 5.1) was performed before IMAC. Buffer containing 50 mM phosphate (pH 8.0) and 300 mM NaCl was used to equilibrate the column. Protein was eluted with a gradient of 1 to 140 mM imidazole.

7.4.5. Immobilised glutathione affinity chromatography

Proteins with GST solubility tag, including *Hpy*CM, *Pfu*CM, and ^{Gsp}CM-^{Tma}DAH7PS, were purified using immobilised glutathione affinity chromatography (IGAC). The 5 mL GSTrap^{TM} HP columns (GE Healthcare) were used for IGAC purifications following same procedures as that of IMAC. Eluted fractions containing the protein of interest might be collected and reloaded on the GSTrap column multiple times to achieve desired purity. Binding buffer was standard phosphate-buffered saline (PBS, 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3), and elution buffer contained 10 mM reduced glutathione and 50 mM Tris-Cl (pH 8.0).

7.4.6. TEV protease treatment

Tobacco etch virus protease (TEV) was used when cleavage the affinity tags off the target protein was required. TEV was expressed and purified from *E. coli* BL21*(DE3) cells containing pRIL and pRK793 plasmids using the IMAC procedure outlined above, and stored at ~1 mg·mL^{-1} in 1 mL aliquots in buffer containing 25 mM KH_2PO_4 (pH 8.0), 200 mM KCl, 10 mM DTT, 2 mM EDTA and 10% (v/v) glycerol. Concentration ratio of 1:100 between TEV and target protein was generally used. The reaction was incubated at 4 °C overnight or until the tag was sufficiently cleaved. After tag removal, the protein was subjected to another IMAC or IGAC step to remove the cleaved tag and the TEV protease.

7.4.7. Size exclusion chromatography

A size exclusion chromatography (SEC) step was used for all protein purifications as a final step to produce highly pure proteins. A HiLoad™ 26/60 Superdex™ 200 prep grade column or a HiLoad™ 16/60 Superdex™ 75 prep grade column (GE Healthcare) was used based on size of the target protein. SEC was performed at 4 °C using buffer containing 0.2 M Tris-Cl (pH 6.0) for mTG, or 10 mM BTP (pH 7.5), 150 mM NaCl (or 100 mM KCl), 200 µM PEP, and 200 µM TCEP for other proteins.

7.4.8. Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SDS-PAGE was performed using Bolt® 10% Bis-Tris Plus protein gels (Invitrogen) in MES running buffer. Protein samples were prepared and denatured following the product manual by adding LDS sample buffer, DTT, and heating at 95 °C for 10 minutes. A premixed Novex® Sharp prestained protein standards (Invitrogen) was used as the molecular weight marker. Electrophoresis was performed at 165 V for 35 minutes and the resulting gels were stained using a heated solution containing 0.1% (w/v) Coomassie brilliant blue R-250, 10% (v/v) glacial acetic acid, and 40% (v/v) methanol for 30 minutes with gentle shaking. The staining step was repeated with fresh staining solution when necessary to generate sufficiently stained gels. Gels were then destained using a heated solution containing 10% (v/v) glacial acetic acid, 40% (v/v) methanol until the protein bands were sufficiently visible.

7.4.9. Determination of protein concentration

Nanodrop® ND-1000 spectrophotometer was used to determine protein concentrations by measuring absorbance at 280 nm with protein-specific molar extinction coefficient values (calculated using ProtParam online tool).[262]

For protein with low extinction coefficient values, such as the CMs, Bradford's assay was used to determine protein concentrations.[263] Bovine serum albumin (BSA) stock was used to create six standards at 0, 0.2, 0.4, 0.6, 0.8 and 1 mg·mL^{-1}. Protein solutions containing each BSA standard (20 µL) were mixed with Bradford reagent to a total volume of 1 mL. The absorbance values of the standards were measured in triplicates at 595 nm to create a calibration curve. Target protein was prepared to approximately 0.2, 0.4 and 0.8 mg·mL^{-1} based on Nanodrop reading and mixed with Bradford reagent. The absorbance values were measured following the same methods as for the standards. Concentrations of the target protein was then calculated based on the standard curve generated from the standards.

Protein samples were concentrated to desired concentration using a Vivaspin^{TM} Turbo 15 mL centrifugal concentrator with 10 kDa or 3 kDa molecular weight cut off (Sartorius Stedim Biotech).

7.4.10. Protein verification

Mass spectrometry

The molecular mass of purified proteins was verified using electrospray ionisation with a Micromass LCT Classic, Bruker maXis 3G at the University of Canterbury Chemistry Department. Protein samples (100 µL) were prepared at 1 mg·mL^{-1} in water.

Circular dichroism

Circular dichroism (CD) was used to verify whether the protein of interest was properly folded with the correct secondary structure. CD spectra were recorded using a JASCO J-815 Spectropolarimeter for range of 250 nm to 190 nm at 25 °C using 1 nm data pitch and bandwidth at 1 s response. Protein solution (~0.03 mg·ml^{-1}) prepared in water in a 3 mL quartz

cuvette with 1 cm path length were used for measurements. Background spectra of water were recorded and subtracted. Resulting data were smoothed by performing a five-point moving average.

7.5. Protein characterisation

7.5.1. Kinetic assays

DAH7PS assay

The standard DAH7PS assay used in this study measured the consumption of PEP at 232 nm (ε = 2.8 x 10^3 $M^{-1}\cdot cm^{-1}$) following published procedures using a Varian Cary 100 UV-visible spectrophotometer.[63, 78, 177] Initial velocity values were determined with respect to varying concentrations of each substrate, when the concentration of the other substrate was fixed at saturating concentrations (defined as at least 5-fold higher than the K_M value). The reaction mixture contained ~0.05 µM protein, specified or saturating concentrations of PEP (for determination of K_M^{E4P}), specified or saturating concentrations of E4P (for determination of K_M^{PEP}), 100 µM of appropriate divalent metal ion in 50 mM BTP buffer (pH 7.5 or as specified). Reactions were initiated by addition of E4P. Assays at elevated temperatures were performed following the same methods. Enzyme activities are specified in U (1 U = consumption of 1 µmol substrate·min^{-1}). Specific enzyme activities are given as U·mg^{-1}. Kinetic parameters were determined by fitting triplicate data to the Michaelis-Menten equation using GraphPad Prism 7.

Specifically, concentration of PEP was fixed at 215 µM when K_M for E4P of *Tma*DAH7PS variants was determined in Chapter 2. Concentration of E4P was fixed at 310 µM when K_M for PEP was determined. For *Tye*DAH7PS, concentration of PEP was fixed at 300 µM when K_M for E4P was determined. Concentration of E4P was fixed at 210 µM when K_M for PEP was

determined. For DAH7PS assays described in Chapter 3, concentration of E4P was varied between 5 µM and 400 µM while PEP was held at 295 µM. Concentration of PEP was varied between 6 µM and 350 µM while E4P was held at 310 µM. Mn^{2+} was used for proteins containing the catalytic barrel from *T. maritima*, whereas Cd^{2+} was used for proteins containing the catalytic barrel from *G. sp.* For *Hpy*DAH7PS described in Chapter 4, concentration of PEP was fixed at 98 µM when K_M for E4P was determined. Concentration of E4P was fixed at 211 µM when K_M for PEP was determined.

For assays assessing the effect of temperature on specific activities of *Tma*DAH7PS and *Tye*DAH7PS variants (Chapter 2), each reaction contained 283 µM PEP, 308 µM E4P, and 100 µM Mn^{2+} in 50 mM BTP with pH fixed at 7.5 at all temperatures tested (30 °C-80 °C). Activity profiles of the chimeric proteins and parent proteins (Chapter 3) at elevated temperatures were assessed with each reaction containing 100 µM appropriate metal ion, 215 µM PEP and 227 µM E4P in 50 mM BTP buffer at pH 7.5 at the temperatures tested (30 °C-80 °C).

Metal dependency studies were performed using the DAH7PS assay with a range of alternative metal ions or EDTA at 100 µM. For inhibition studies, the rate of reaction was measured following incubation of the protein for 5-10 minutes with varying concentrations of Tyr or Phe (0 to 1 mM), or 3 to 5 minutes-incubation with varying concentrations of prephenate (0 to 500 µM). For assays containing mixture of two proteins of interest, such as *Hpy*CM and *Hpy*DAH7PS, the reaction mixture was incubated for at least 5 minutes before the reaction was initiated.

CM activity

Standard assays for the CM-catalysed reactions were measured by monitoring consumption of chorismate at 274 nm (ε = 2.630 x 10^3 $M^{-1}\cdot cm^{-1}$).[30, 78] Buffers used were same as that for

DAH7PS assays and reactions were initiated by addition of chorismate (9 µM to 400 µM). *Hpy*CM activity was measured at 30 °C. Activities of *Pfu*CM, ^{Gsp}CM-^{Tma}DAH7PS and *Gsp*DAH7PS variants were measured at 50 °C.

Chorismic acid (CA) stock solutions used for kinetic assays were prepared in appropriate buffer at an approximate concentration of 10 mM. Accurate concentrations of CA were determined using a standard CM assay. The reaction mixture contained 0.05 µM enzyme, 50 mM BTP (pH 7.5) and 5 µL of CA stock with total volume of 1 mL. Based on Beer's law ($A = \varepsilon \cdot L \cdot C$), the change in absorbance from before enzyme was added until the reaction had reached completion was used to calculate the change in concentration of CA (ε = 2.630 x 10^3 $M^{-1}cm^{-1}$) upon addition of enzyme. The change in absorbance did not account for the increase in background absorbance due to addition of enzyme, so correction factor was determined by an identical assay without CA present and was subtracted to calculate the corrected concentrations. The concentration of PEP and E4P were determined using similar methods.

7.5.2. Differential scanning fluorimetry

Melting temperatures of the proteins were determined by using differential scanning fluorimetry with an iCycler iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad). Protein samples (0.1 $mg \cdot mL^{-1}$) were prepared in storage SEC buffer with SYPRO orange dye in the presence of appropriate metal ions (100 µM) and ligands (1 mM Tyr/Phe/Trp, or 500 µM prephenate). Controls were similarly prepared for each condition, with buffer in place of protein. The temperature was set to increase from 20-98 °C at a rate of 1°C per minute and fluorescence was measured in 0.2 °C increments. Melting temperatures were determined as the temperature of maximum inflection of the first melting event after subtracting background from the control. Errors were calculated based on standard deviation from triplicate measurements.

7.5.3. Determination of protein oligomeric state

Analytical size exclusion chromatography and analytical ultracentrifugation (AUC) were typically used to determine the size of the target protein in solution. This information was also derived from SAXS experiments when required.

Analytical SEC

A Superdex^{TM} 200 10/300 GL column (GE Healthcare) was used for analytical SEC. It was equilibrated with SEC buffer and calibrated by protein standards with known molecular weights, including ovalbumin, conalbumin, aldolase, ferritin, thyroglobulin and blue dextran. Protein standards and samples were applied to the column at 1 mg·mL^{-1} in 500 µL volume unless otherwise stated. The elution volume of blue dextran was recorded as the void volume. Elution volumes for all samples were recorded and graphed against the log [protein standard mass (Da)] producing a linear graph, from which unknown protein masses were calculated.

AUC

Sedimentation velocity (SV) and sedimentation equilibrium (SE) experiments were performed using a Beckman Coulter Model XL-I analytical ultracentrifuge equipped with UV/Vis scanning optics. Reference buffer solution (20 mM Tris–HCl or BTP, 150 mM NaCl, pH 7.5) and sample solutions were loaded into 12 mm double-sector cells with quartz windows and the cells were then mounted in an An-60 Ti 4-hole rotor.

Purified and dialysed protein samples at concentrations of 0.1 to 1.6 mg·mL^{-1} and the reference buffer were centrifuged at 50000 rpm (for SV) at 20 °C and absorbance data were collected in continuous mode at 278 nm without averaging. Data were fitted to a continuous size distribution model using the program SEDFIT.[129] The SE data were collected over a concentration gradient at 7000, 10000, 12000, and 18000 rpm at 20 °C and 280 nm. The partial

specific volume of the sample, buffer density and viscosity were calculated using the program SEDNTERP.[130]

For analysis of the SE data collected from *Tye*DAH7PS, the meniscus, cell bottom, and fitting limits were adjusted using SEDFIT. The data was then exported for further processing with SEDNTREP. The 'monomer-dimer self-association' model was selected as the protein was expected to display dimer-tetramer equilibrium, and the Global Parameters window was then completed by filling the molecular weight and sedimentation coefficient of each component. Global Fit was initiated until the fit converged. Fit qualities from various models were compared based on the reduced global χ^2 values as well as fitting residuals to identify the best-fitted model. Key information such as log *Ka*, χ^2, and errors were obtained under the Statistics function.

7.5.4. Small angle X-ray scattering

Small angle X-ray scattering (SAXS) measurements were collected at the Australian Synchrotron SAXS/WAXS beamline equipped with a Pilatus detector (1 M, 170 × 170 mm, effective pixel size, 172 × 172 μm). The X-ray wavelength was 1.0332 Å. The sample-detector distance was 1.6 m, which delivered a *s* range of 0.01–0.40 $Å^{-1}$ [where *s* is the magnitude of the scattering vector, which is related to the scattering angle (2θ) and the wavelength (λ) as follows: $s = (4\pi/\lambda)\cdot\sin\theta$].

For SEC samples, scattering data were collected at 25 °C (unless otherwise specified) following elution of the protein samples (< 10 $mg\cdot mL^{-1}$) from a size-exclusion chromatography column (Superdex 200_Increase 5/150), which was pre-equilibrated with buffer containing 10 mM BTP (pH 7.5), 100 mM KCl, 200 μM PEP, and 3% (v/v) glycerol in the absence or presence of inhibitors (1 mM Tyr or 500 μM prephenate). For static samples, proteins were prepared

with serial dilutions at concentration range of 0.5 to 4 mg·mL^{-1} and filtered before loading in 96-well plates. Scattering data were collected directly from the degassed samples in the plates.

Raw data were processed and background-subtracted using Scatterbrain.[63] Scatterings from $A280_{nm}$ peaks were summed and averaged for SEC experiments. All scatterings excluding outliers were summed and averaged for static experiments. Guinier plots were linear for $sR_g < 1.3$. Scattering intensity (*I*) versus *s* of each protein was generated with Primus and plots were scaled for comparison.[264] Theoretical scattering profiles were generated from model coordinates were compared and fitted with corresponding experimental scatterings with Crysol.[187] Specifically, the model coordinates used in this study include crystal structures of *Tma*DAH7PS PDB 1RZM (apo), 3PG9 (Tyr), homology model of *Gsp*DAH7PS (apo) and crystal structure 5J6F (prephenate), crystal structure of *Mtu*DAH7PS (3NV8), homology model of *Hpy*DAH7PS, and crystal structure of *Hpy*CM (6AL9).

7.5.5. X-ray crystallography

Crystallisation

Automated screening for protein crystals was performed using Mosquito® Crystal robot (TTP Labtech) with sitting drop vapour diffusion technique. The screen conditions typically included PACT premier™ HT-96, JCSG-plus™ HT-96, Clear Strategy™ I HT-96, Clear Strategy™ II HT-96, ProPlex™ HT-96, MIDASplus™ HT-96, and Morpheus® II HT-96 crystallisation screens (Molecular Dimensions). The protein samples (400 nL) were diluted to half by addition of a screening condition (400 nL). Volume of 40 µL of each screening solutions were used as reservoir. When required, ligands with appropriate concentrations were added into the sample drops and co-crystallised. Manual screening using 10 mL reagent kits was also occasionally performed with 24-well hanging drop plates. Plates were incubated at 20 or 5 °C and examined regularly to record crystal growth.

If a suitable lead condition was identified, the condition was generally optimised to improve crystal growth. The optimisations were performed using hanging drop vapour diffusion technique. Conditions such as pH, salt and precipitant concentrations, and protein concentrations were varied to achieve optimal formation of crystals. Sample drop size was 2 or 4 µL with 500 µL reservoir solution in a 24-well hanging drop plates (Hampton Research).

As described in Chapter 2.3, leading conditions for producing *Tma*DAH7PS^{S31I} crystals in the presence of Phe were optimised. These conditions are listed below.

PACT C7 0.2 M Sodium chloride, 0.1 M HEPES (pH 7.0), 20% w/v PEG 6000

PACT G5 0.2 M Sodium nitrate, 0.1 M Bis-Tris propane (pH 7.5), 20% w/v PEG 3350

JCSG+ B8 0.2 M Magnesium chloride hexahydrate, 0.1 M Tris (pH7.0), 10% w/v PEG 8000

Conditions that produced ideal crystals of *Hpy*CM were optimised from MIDAS conditions D8 and G7 as listed below. Chorismate at a final concentration of 500 µM was included in all conditions. Conditions optimised from MIDAS G7 produced crystals with optimal quality. A relatively long incubation at 20 °C for at least three days was required for crystal formation.

MIDAS D8 0.1 M Tris (pH 8.0), 20% v/v glycerol ethoxylate

MIDAS G7 0.2 M Ammonium acetate, 0.1 M MES (pH 6.5), 30% v/v glycerol ethoxylate

Structure determination and refinement

X-ray diffraction data of the protein crystals were collected at the Australian Synchrotron using Macromolecular Crystallography (MX1) or Micro Crystallography (MX2) beamline. Detailed

structure determination and refinement of *Hpy*CM can be found in Chapter 4.3.3. Briefly, data were processed using X-ray Detector Software (XDS), AIMLESS and TRUNCATE (CCP4 program suite).[123-125] Initial phase was predicted by molecular replacement using MOLREP with coordinates from crystal structures or models as search models.[126] CHAINSAW was used when smaller units of the search models were required.[201] Density Modification (CCP4 program suite) was used when improvement of density was required. If a satisfactory solution was identified, multiple refinement processes were performed using rigid body, NCS, TLS and restrained refinements using Refmac5 with iterative model building using Coot.[203, 265] Water molecules and ligands were added at last by interpretation of the 2|Fo|-|Fc| and |Fo|-|Fc| maps. Resulting structure models were validated and adjusted using the Coot validation tools and wwPDB deposition servers before deposition.

Appendix

Appendix A - Supporting information for Chapter 1

Mechanism of the CM-catalysed reaction

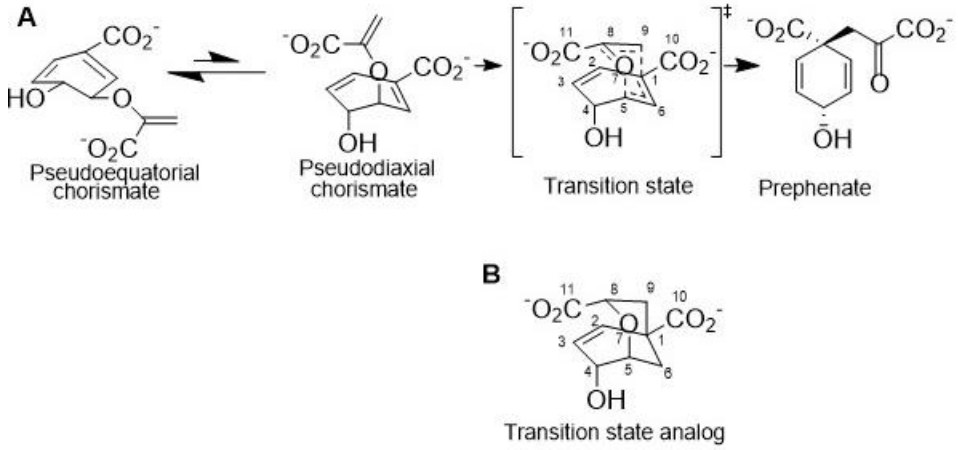


Figure S1. Claisen rearrangement of chorismate to prephenate. The two conformers of chorismate are shown. Reaction occurs via a chair-like transition state (atom numbers labeled, adapted from Walsh *et al*.[27]). B. Transition state analog, *endo*-oxabicyclic dicarboxylic acid. Bond formation and cleavage are shown as dashed lines.

The CM-catalysed reaction is a pericyclic Claisen rearrangement involving an ether oxygen bond cleavage between C5 and O7, followed by the formation of a new bond between C1 and C9 (Figure S1). In solution, the chorismate molecule favours a lower energy pseudoequatorial conformation, which is in equilibrium with the pseudodiaxial conformation. However, CM selects the minor pseudodiaxial conformer, which is set up for the rearrangement. Both uncatalysed and catalysed rearrangements of chorismate to prephenate occur via a chair-like

transition state.[27] The *endo*-oxabicyclic dicarboxylic acid (Figure S1 B), a geometric mimic of this transition state, is currently the best inhibitor of a broad range of CM, which is consistent with the idea that the enzyme catalysed reaction proceeds through a corresponding pericyclic transition structure.[266] CM accelerates the rate of reaction by 2×10^6 fold compared to the uncatalysed reaction. The origins of this remarkable rate enhancement still remain a matter of heated debates.[267-273]

Appendix B - Supporting information for Chapter 2

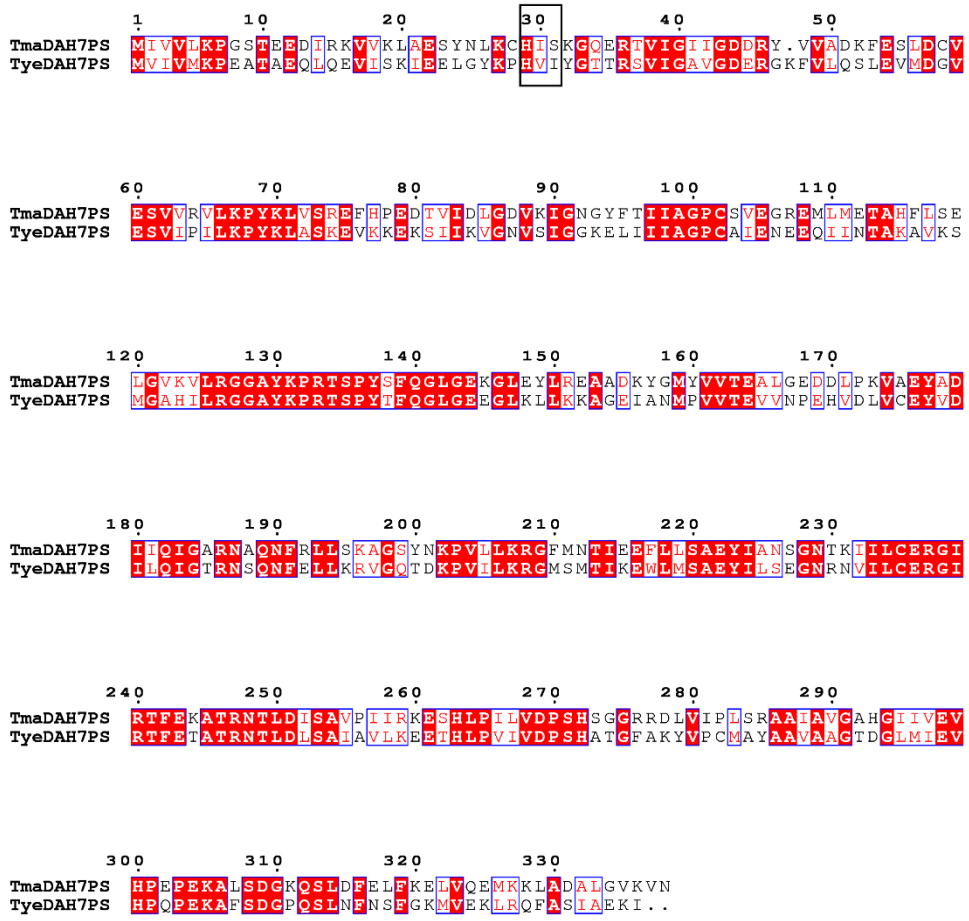


Figure S2. Sequence alignment of *Tma*DAH7PS and *Tye*DAH7PS. Numbering corresponds to *Tma*DAH7PS. The region of interest is highlighted with black rectangle. Sequences were aligned using ClustalW and graphed using ESPript.[257, 274]

Table S1. Molecular mass confirmation of *Tma*DAH7PS and *Tye*DAH7PS variants. Calculated molecular weight is compared with experimentally determined value from mass spectrometry.

Protein	Theoretical mass (Da)	Measured mass (Da)
*Tma*DAH7PS^{WT}	37378.2	37379.4
*Tma*DAH7PS^{S31I}	37404.2	37405.1
*Tma*DAH7PS^{S31V}	37390.2	37390.5
*Tye*DAH7PS^{WT}	37989.1	37992.4
*Tye*DAH7PS^{I31S}	37963.1	37965.9

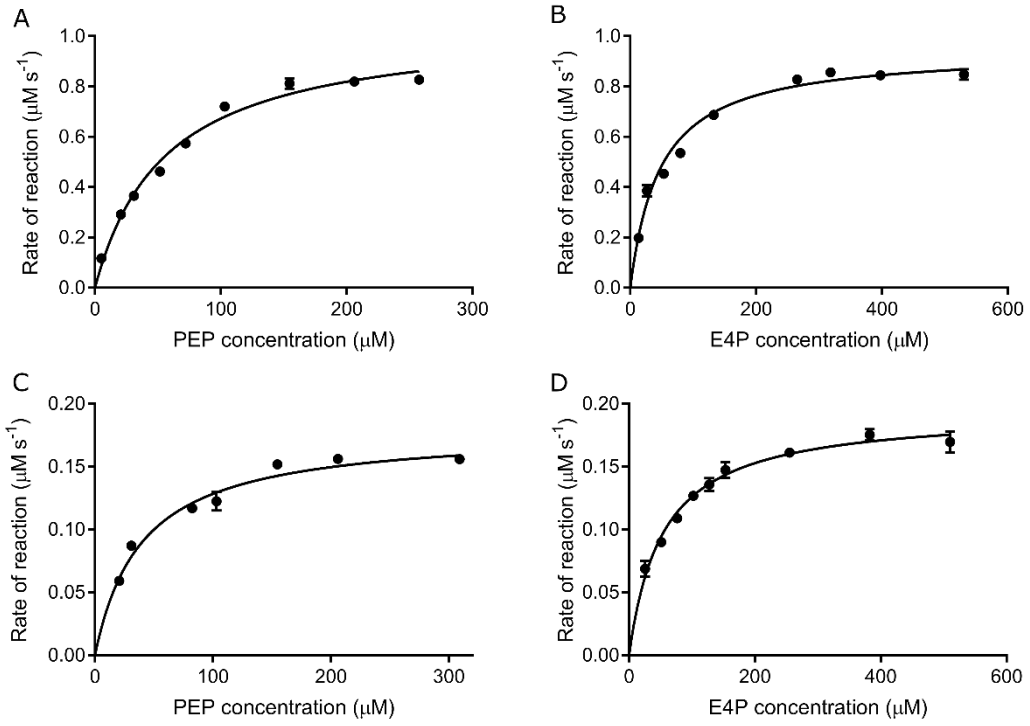


Figure S3. Michaelis–Menten kinetics of *Tma*DAH7PS^{S31I} (A, B) and *Tye*DAH7PS^{I31S} (C, D).

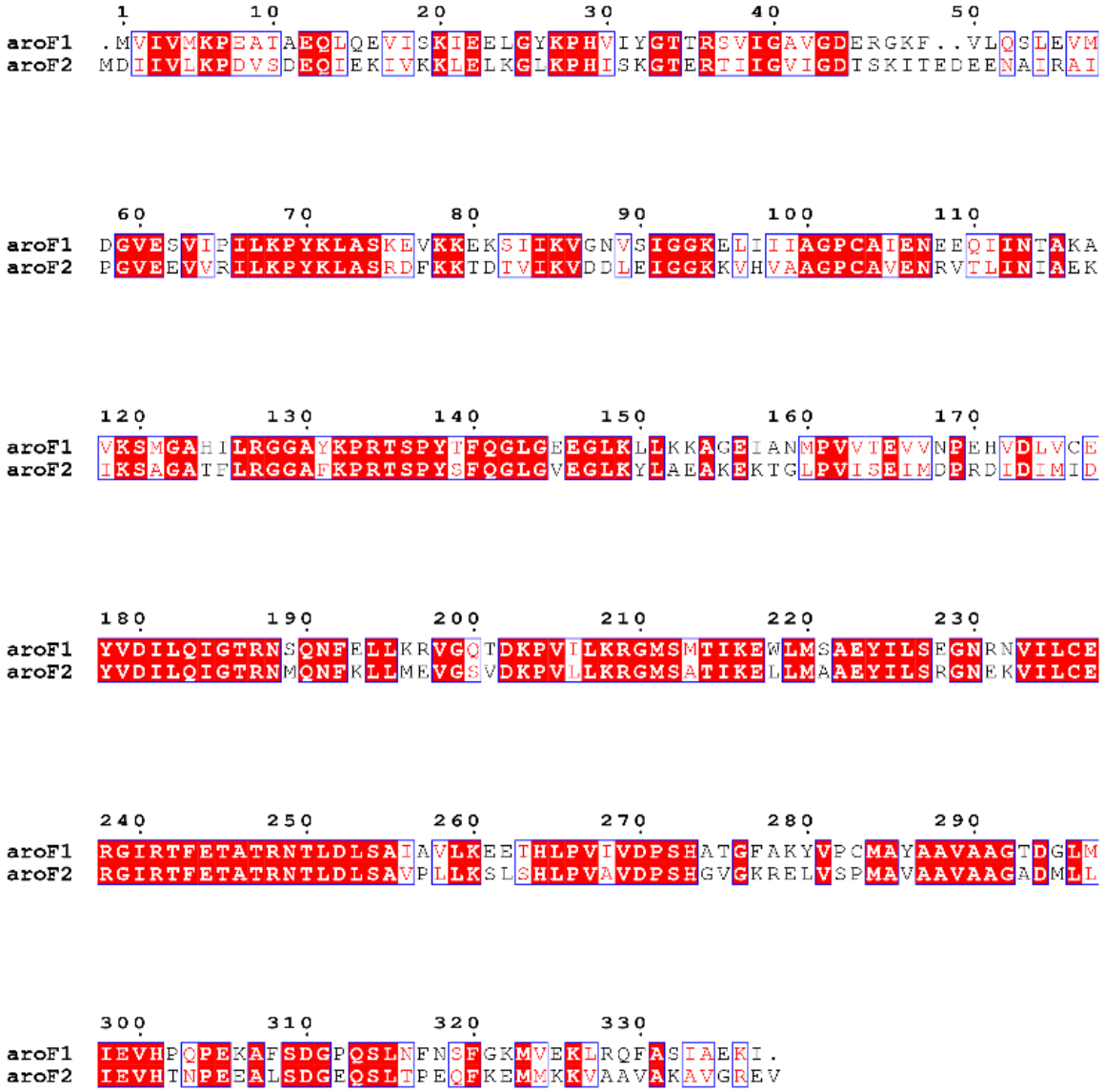


Figure S4. Sequence alignment of the two Type Iβ DAH7PS enzymes in *T. yellowstonii*. The *aroF1* DAH7PS is characterised in this study.

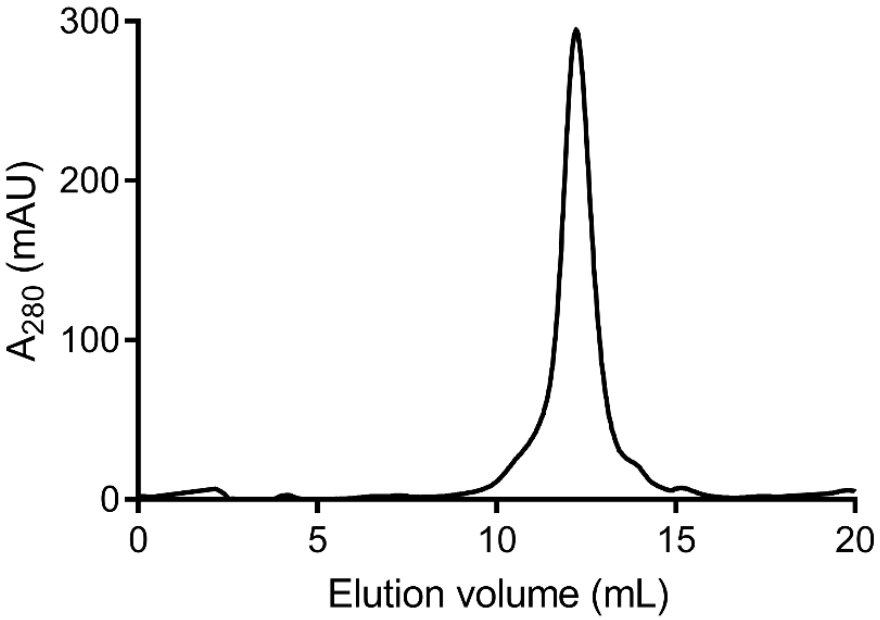


Figure S5. Gel filtration of N-terminal His_6-*Tye*DAH7PS.* This protein is solely tetrameric.

*Three constructs of *Tye*DAH7PS (listed below) were created in this study as mentioned in Chapter 2. Primers used for these constructs can be found in Table 7.1.

1. C-terminal His_6-tagged construct was used in all experiments.
2. Construct without purification tag was used as a control.
3. N-terminal His_6-tagged construct created solely tetrameric protein.

Appendix C - Supporting information for Chapter 3

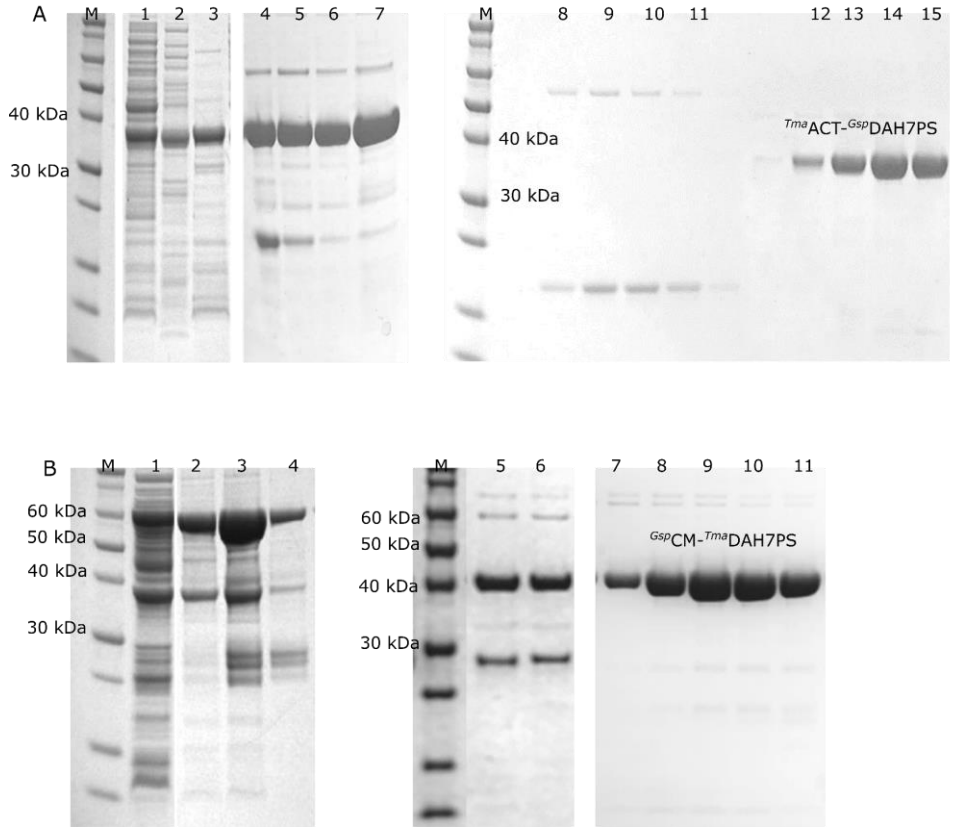


Figure S6. Purification of *Tma*ACT-*Gsp*DAH7PS (A) and *Gsp*CM-*Tma*DAH7PS (B). A. 1, soluble fraction after cell lysis and centrifugation. 2, insoluble fraction. 3, soluble fraction after heat treatment. 4-7, eluted fractions from HIC. 8-11, impurities eluted from SEC. 12-15, purified *Tma*ACT-*Gsp*DAH7PS from SEC. B. 1, soluble fraction after cell lysis and centrifugation. 2-4, eluted fractions from GSTrap. 5-6, protein after TEV treatment with the higher band corresponding to the protein of interest, and lowered band corresponding to cleaved GST tag. 7-11, purified *Gsp*CM-*Tma*DAH7PS from SEC. M, protein marker.

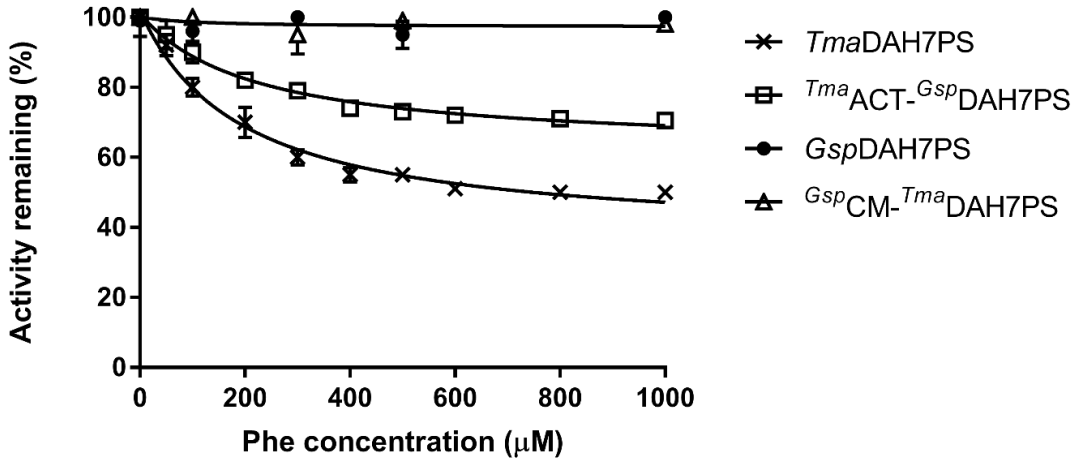


Figure S7. DAH7PS activity of the parent and chimeric proteins in the presence of 1 mM Phe. Wild-type *Tma*DAH7PS and ^{Tma}ACT-^{Gsp}DAH7PS showed inhibition with Phe.

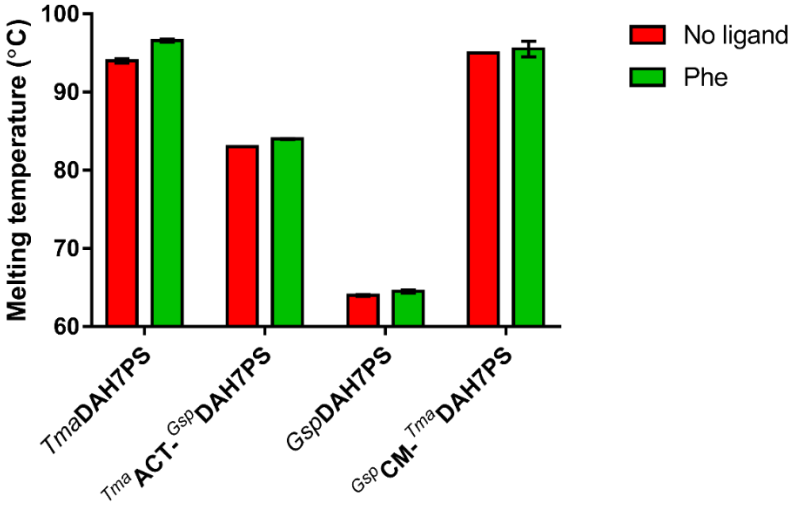


Figure S8. Thermostability of the parent and chimeric proteins in the presence or absence of 1 mM Phe. Error bars represent SD of triplicate measurements.

Appendix D - Supporting information for Chapter 4

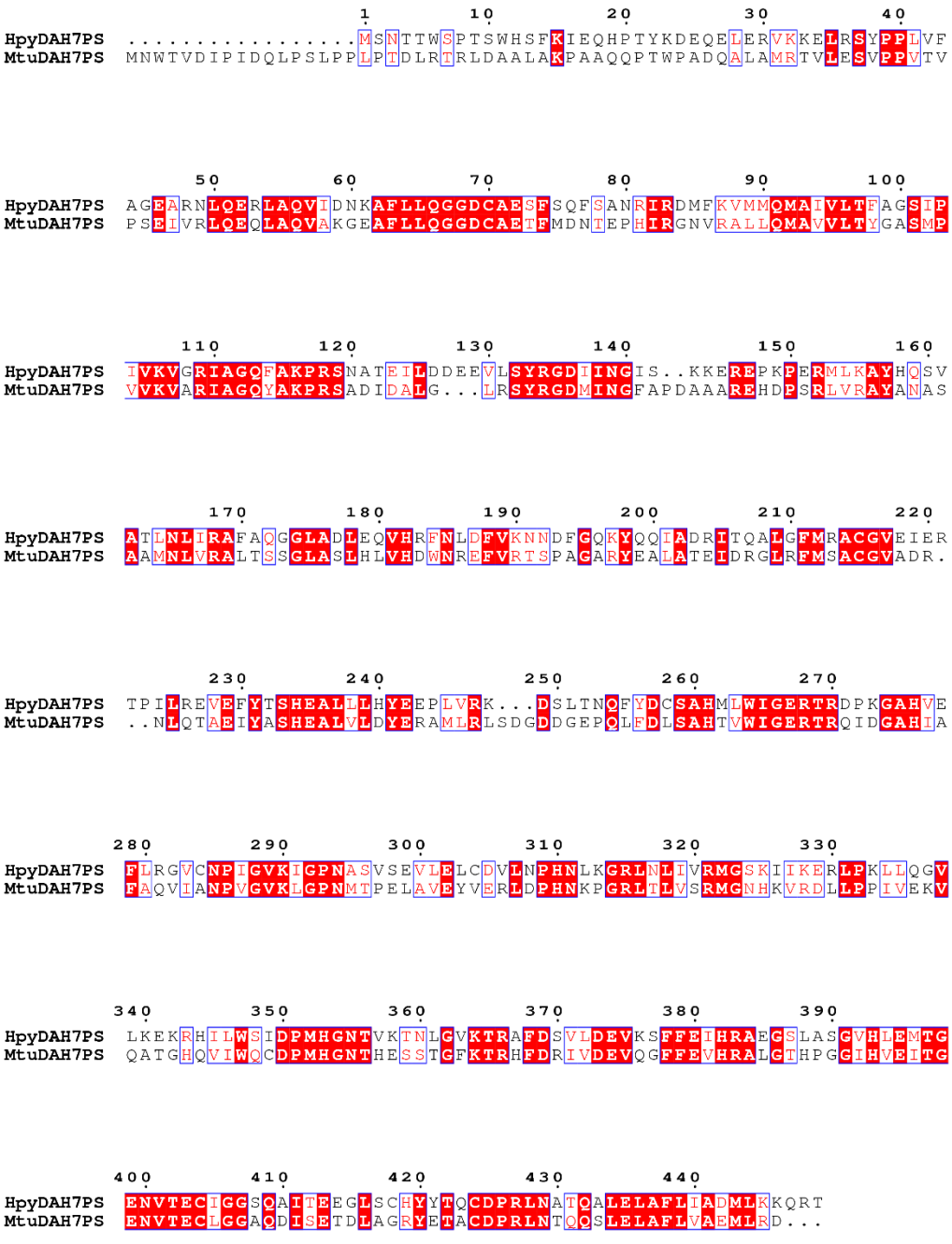


Figure S9. Sequence alignment of *Hpy*DAH7PS and *Mtu*DAH7PS generated by ClustalW and graphed by ESPript.[257, 274]

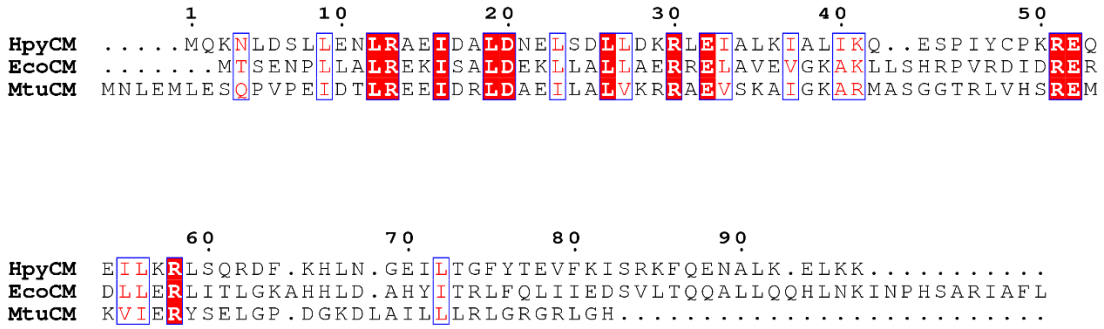


Figure S10. Sequence alignment of *Hpy*CM, *Eco*CM, and *Mtu*CM generated by ClustalW and graphed by ESPript.[257, 274]

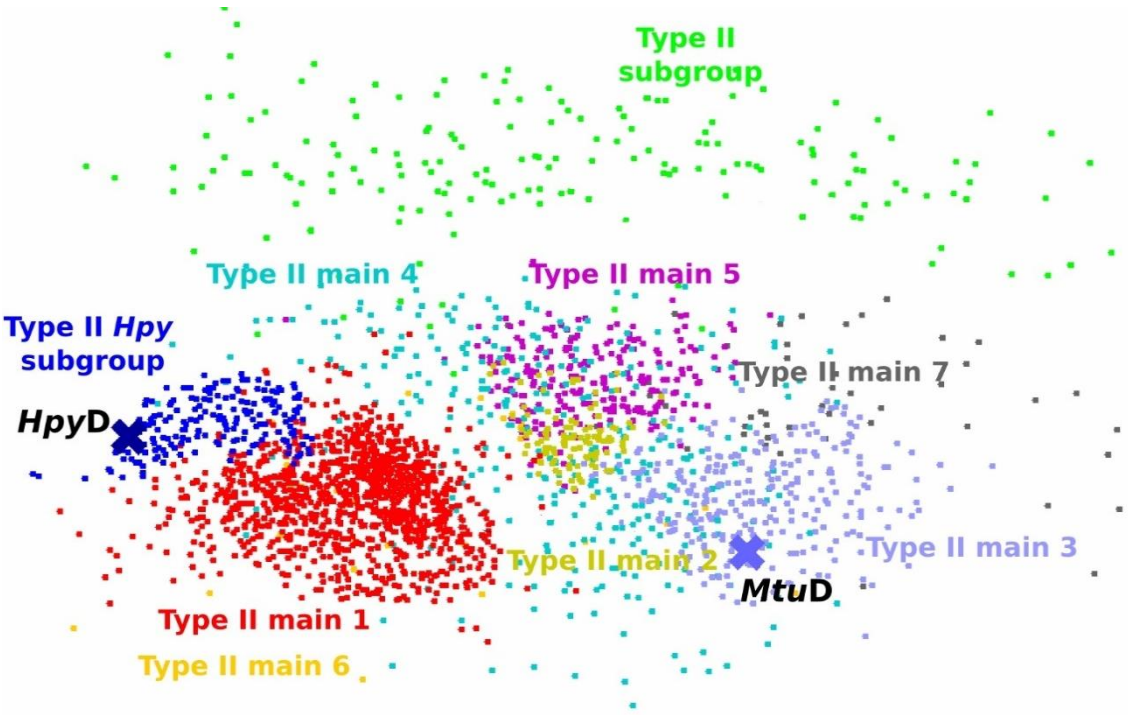


Figure S11. CLANS clustering of Type II DAH7PS sequences. Each dot represents a sequence. *Hpy*DAH7PS and *Mtu*DAH7PS are labelled. Figure was adapted from Lang.[255]

Appendix E - Supporting information for Chapter 5

Table S2. Kinetic parameters of various enzymes used for protein labelling. Table was adapted from Rashidian *et al.*[218]

Enzyme	Peptide tag	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M (μM^{-1} min^{-1})	Site of the tag
Formylglycine generating enzyme	CXPXR or 13-mer LCTPSRGSLFTGR	-	-	-	N-/C-terminus
Phosphopantetheinyl transferase	ACP, PCP or ybbR tags (11 mer: DSLEFIASKLA; 13mer: VLDSLEFIASKLA; 17 mer: GSQDVLDSLEFIASKLA)	60.8	14.7	0.242[a]	N-/C-terminus or loops
Sortase	LPXTG	5,500	16.2	0.003	Any site
Transglutaminase	XXQXX	7	45	6	Any site
Farnesyltransferase	CaaX	1.71	31.2	18.2[b]	C-terminus
Biotin ligase	GLNDIFEAQKIEWHE	4.2	9.6	2.3c	N-/C-terminus
Lipoic acid ligase	GFEIDKVWYDLDA	2.88	-	-	N-/C-terminus or loops

[a]500 μM ybbR 13 mer and biotin-CoA;

[b]yPFTase and 2.4 μM CVIA for FPP;

[c]*E.coli* biotin ligase and biotin

Table S3. Molecular mass confirmation of *Tma*DAH7PS and *Gsp*DAH7PS variants. Calculated molecular weight is compared with experimentally determined value from mass spectrometry.

Protein	Theoretical mass (Da)	Measured mass (Da)
*Tma*DAH7PS^{C58S}	37362.1	37362.2
*Tma*DAH7PS$^{C58S\text{-}C235S}$	37346.0	37346.5
*Gsp*DAH7PS^{WT}	41374.4	41374.8
*Gsp*DAH7PS^{D302C}	41362.5	41362.2
*Gsp*DAH7PS^{C307S}	41358.3	41358.1
*Gsp*DAH7PS$^{C307S\text{-}D302C}$	41346.4	42346.4

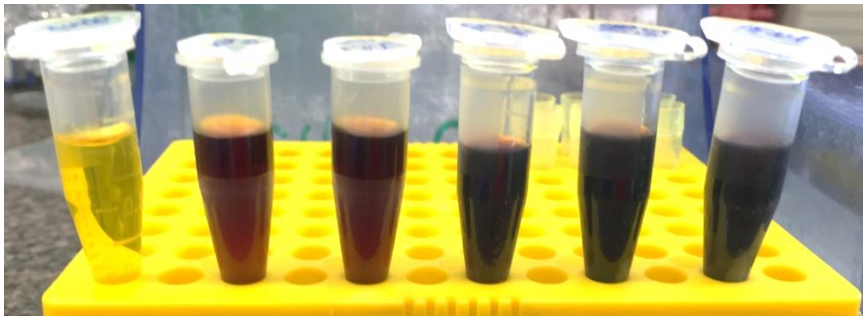


Figure S12. Activity test for mTG. Iron complex produced after quenching of the mTG-catalysed reaction. Each tube contained increasing concentration of mTG from left (blank) to right.

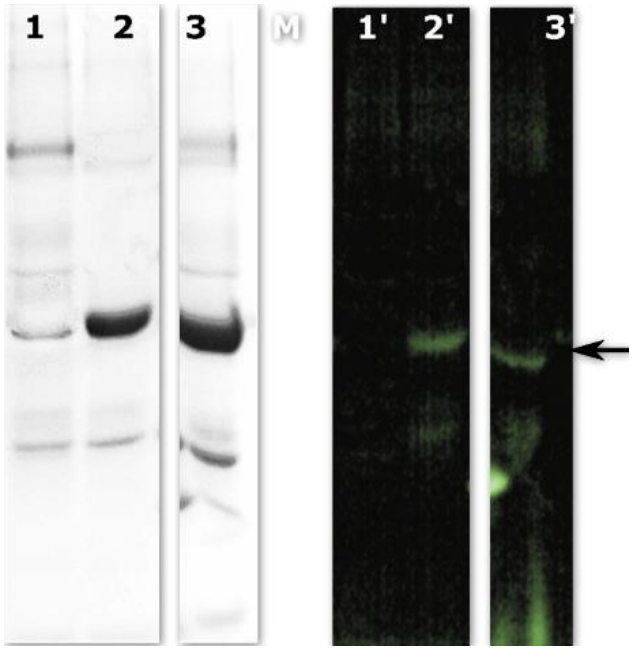


Figure S13. DNS-Cd labelling of *Gsp*DAH7PS. Labelled protein largely remained in insoluble fraction. 1. Soluble fraction after centrifugation. 2. Insoluble fraction after centrifugation. 3. Crude mixture before centrifugation. Arrow indicates the size of the target protein.

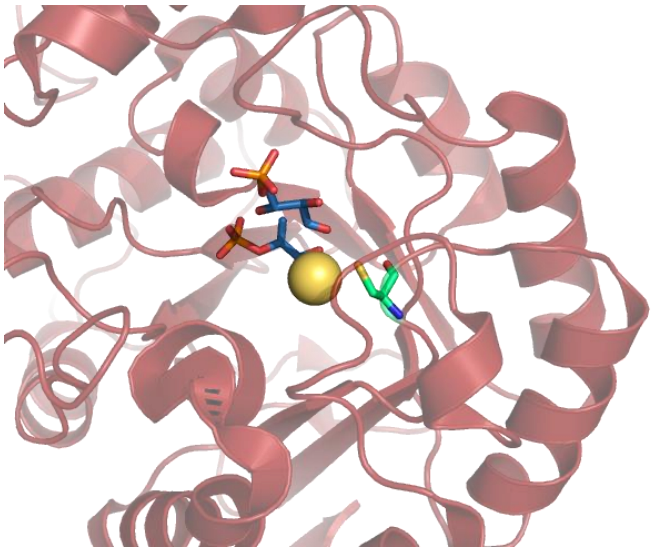


Figure S14. Active site of *Tma*DAH7PS with PEP, E4P and Cd^{2+}. Substrates are shown as blue sticks. Cys102 is highlighted as green sticks. Cd^{2+} is shown as yellow sphere. Figure produced from PDB 1RZM.

References

1. Hartwell, L. H.; Hopfield, J. J.; Leibler, S.; Murray, A. W., From molecular to modular cell biology. *Nature* **1999,** *402*, C47-C52.

2. Monod, J.; Wyman, J.; Changeux, J. P., On the nature of allosteric transitions: a plausible model. *J Mol Biol* **1965,** *12*, 88-118.

3. Koshland, D. E.; Némethy, G.; Filmer, D., Comparison of experimental binding data and theoretical models in proteins containing subunits*. *Biochemistry* **1966,** *5* (1), 365-385.

4. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **1958,** *181* (4610), 662-666.

5. Pardee, A. B.; Reddy, G. P., Beginnings of feedback inhibition, allostery, and multi-protein complexes. *Gene* **2003,** *321*, 17-23.

6. Novick, A.; Szilard, L., Experiments with the chemostat on the rates of amino acid synthesis in bacteria. *Dyn Growth Process* **1954**, 21-32.

7. Cui, Q.; Karplus, M., Allostery and cooperativity revisited. *Protein Sci* **2008,** *17* (8), 1295-1307.

8. Whitley, M. J.; Lee, A. L., Frameworks for understanding long-range intra-protein communication. *Curr Protein Pept Sci* **2009,** *10* (2), 116-127.

9. Changeux, J. P., Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu Rev Biophys* **2012,** *41*, 103-133.

10. Cooper, A.; Dryden, D. T., Allostery without conformational change. A plausible model. *Eur Biophys J* **1984,** *11* (2), 103-109.

11. Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G., Dynamically driven protein allostery. *Nat Struct Mol Biol* **2006,** *13* (9), 831-838.

12. Tzeng, S.-R.; Kalodimos, C. G., Dynamic activation of an allosteric regulatory protein. *Nature* **2009,** *462* (7271), 368-372.

13. Tzeng, S. R.; Kalodimos, C. G., Protein activity regulation by conformational entropy. *Nature* **2012,** *488* (7410), 236-240.

14. Petit, C. M.; Zhang, J.; Sapienza, P. J.; Fuentes, E. J.; Lee, A. L., Hidden dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A* **2009,** *106* (43), 18249-18254.

15. Perutz, M. F., Stereochemistry of cooperative effects in haemoglobin: haem-haem interaction and the problem of allostery. *Nature* **1970,** *228* (5273), 726-734.

16. Perutz, M. F.; Wilkinson, A. J.; Paoli, M.; Dodson, G. G., The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Annu Rev Biophys Biomol Struct* **1998,** *27*, 1-34.

17. Sprenger, G. A., Aromatic amino acids. In *Amino Acid Biosynthesis - Pathways, Regulation and Metabolic Engineering*, Wendisch, V. F., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; 93-127.

18. Vogt, T., Phenylpropanoid biosynthesis. *Mol Plant* **2010,** *3* (1), 2-20.

19. Razal, R. A.; Ellis, S.; Singh, S.; Lewis, N. G.; Towers, G. N., Nitrogen recycling in phenylpropanoid metabolism. *Phytochemistry* **1996,** *41* (1), 31-35.

20. Maeda, H.; Dudareva, N., The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu Rev Plant Biol* **2012,** *63*, 73-105.

21. Kutchan, T. M., Alkaloid biosynthesis-the basis for metabolic engineering of medicinal plants. *Plant Cell* **1995,** *7* (7), 1059-1070.

22. Radwanski, E. R.; Last, R. L., Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. *Plant Cell* **1995,** *7* (7), 921-934.

23. Herrmann, K. M.; Weaver, L. M., The shikimate pathway. *Annu Rev Plant Biol* **1999,** *50* (1), 473-503.

24. Bentley, R., The shikimate pathway: a metabolic tree with many branches. *Crit Rev Biochem Mol Biol* **1990,** *25*, 307 - 384.

25. Fernstrom, J. D.; Fernstrom, M. H., Tyrosine, phenylalanine, and catecholamine synthesis and function in the brain. *J Nutr* **2007,** *137* (6 Suppl 1), 1539S-1547S; discussion 1548S.

26. Herrmann, K. M., The shikimate pathway as an entry to aromatic secondary metabolism. *Plant Physiol* **1995,** *107* (1), 7-12.

27. Walsh, C. T.; Liu, J.; Rusnak, F.; Sakaitani, M., Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. *Chem Rev* **1990,** *90* (7), 1105-1129.

28. Goyal, S.; Yuan, J.; Chen, T.; Rabinowitz, J. D.; Wingreen, N. S., Achieving optimal growth through product feedback inhibition in metabolism. *PLOS Comp Biol* **2010,** *6* (6), e1000802.

29. Krappmann, S.; Lipscomb, W. N.; Braus, G. H., Coevolution of transcriptional and allosteric regulation at the chorismate metabolic branch point of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* **2000,** *97* (25), 13585-13590.

30. Sasso, S.; Okvist, M.; Roderer, K.; Gamper, M.; Codoni, G.; Krengel, U.; Kast, P., Structure and function of a complex between chorismate mutase and DAHP synthase: efficiency boost for the junior partner. *EMBO J* **2009,** *28* (14), 2128-2142.

31. Schmidheini, T.; Mosch, H. U.; Evans, J. N.; Braus, G., Yeast allosteric chorismate mutase is locked in the activated state by a single amino acid substitution. *Biochemistry* **1990,** *29* (15), 3660-3668.

32. Nida, D. L.; Kolacz, K. H.; Buehler, R. E.; Deaton, W. R.; Schuler, W. R.; Armstrong, T. A.; Taylor, M. L.; Ebert, C. C.; Rogan, G. J.; Padgette, S. R., Glyphosate-tolerant cotton: genetic characterization and protein expression. *J Agric Food Chem* **1996,** *44* (7), 1960-1966.

33. Kishore, G. M.; Shah, D. M., Amino acid biosynthesis inhibitors as herbicides. *Annu Rev Biochem* **1988,** *57*, 627-263.

34. Coggins, J.; Abell, C.; Evans, L.; Frederickson, M.; Robinson, D.; Roszak, A.; Lapthorn, A., Experiences with the shikimate-pathway enzymes as targets for rational drug design. *Biochem Soc Trans* **2003,** *31* (3), 548-552.

35. Dougan, G.; Chatfield, S.; Pickard, D.; Bester, J.; O'Caliaghan, D.; Maskell, D., Construction and characterization of vaccine strains of *Salmonella* harboring mutations in two different aro genes. *J Infect Dis* **1988,** *158* (6), 1329-1335.

36. Günel-Özcan, A.; Brown, K. A.; Allen, A. G.; Maskell, D. J., *Salmonella typhimurium* aroB mutants are attentuated in BALB/c mice. *Microb Pathog* **1997,** *23* (5), 311-316.

37. Hoiseth, S. K.; Stocker, B., Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature* **1981,** *291*, 239-239.

38. Lowe, D. C.; Savidge, T. C.; Pickard, D.; Eckmann, L.; Kagnoff, M. F.; Dougan, G.; Chatfield, S. N., Characterization of candidate live oral *Salmonella typhi* vaccine strains

harboring defined mutations in aroa, aroc, and htra. *Infect Immun* **1999,** *67* (2), 700-707.

39. Cersini, A.; Salvia, A. M.; Bernardini, M. L., Intracellular multiplication and virulence of *Shigella flexneri* auxotrophic mutants. *Infect Immun* **1998,** *66* (2), 549-557.

40. Noriega, F. R.; Wang, J. Y.; Losonsky, G.; Maneval, D. R.; Hone, D. M.; Levine, M. M., Construction and characterization of attenuated delta aroA delta virG *Shigella flexneri* 2a strain CVD 1203, a prototype live oral vaccine. *Infect Immun* **1994,** *62* (11), 5168-5172.

41. Bowe, F.; O'Gaora, P.; Maskell, D.; Cafferkey, M.; Dougan, G., Virulence, persistence, and immunogenicity of *Yersinia enterocolitica* O: 8 aroA mutants. *Infect Immun* **1989,** *57* (10), 3234-3236.

42. Foulongne, V.; Walravens, K.; Bourg, G.; Boschiroli, M. L.; Godfroid, J.; Ramuz, M.; O'Callaghan, D., Aromatic compound-dependent *Brucella suis* is attenuated in both cultured cells and mouse models. *Infect Immun* **2001,** *69* (1), 547-550.

43. Moral, C. H.; del Castillo, E. F.; Fierro, P. L.; Cortés, A. V.; Castillo, J. A.; Soriano, A. C.; Salazar, M. S.; Peralta, B. R.; Carrasco, G. N., Molecular characterization of the *Aeromonas hydrophila* aroa gene and potential use of an auxotrophicaroa mutant as a live attenuated vaccine. *Infect Immun* **1998,** *66* (5), 1813-1821.

44. McArthur, J. D.; West, N. P.; Cole, J. N.; Jungnitz, H.; Guzmán, C. A.; Chin, J.; Lehrbach, P. R.; Djordjevic, S. P.; Walker, M. J., An aromatic amino acid auxotrophic mutant of *Bordetella bronchiseptica* is attenuated and immunogenic in a mouse model of infection. *FEMS Microbiol Lett* **2003,** *221* (1), 7-16.

45. Srilunchang, T.; Proungvitaya, T.; Wongratanacheewin, S.; Strugnell, R.; Homchampa, P., Construction and characterization of an unmarked aroC deletion mutant of *Burkholderia pseudomallei* strain A2. *Southeast Asian J Trop Med Public Health* **2009,** *40* (1), 123.

46. Jaworski, E. G., Mode of action of N-phosphonomethylglycine. Inhibition of aromatic amino acid biosynthsis. *J Agric Food Chem* **1972,** *20* (6), 1195-1198.

47. Duke, S. O.; Powles, S. B., Glyphosate: a once-in-a-century herbicide. *Pest Manag Sci* **2008,** *64* (4), 319-325.

48. Roberts, F.; Roberts, C. W.; Johnson, J. J.; Kyle, D. E.; Krell, T.; Coggins, J. R.; Coombs, G. H.; Milhous, W. K.; Tzipori, S.; Ferguson, D. J.; Chakrabarti, D.; McLeod, R., Evidence for the shikimate pathway in apicomplexan parasites. *Nature* **1998,** *393* (6687), 801-805.

49. Berry, A., Improving production of aromatic compounds in *Escherichia coli* by metabolic engineering. *Trends Biotechnol* **1996,** *14* (7), 250-256.

50. Bongaerts, J.; Krämer, M.; Müller, U.; Raeven, L.; Wubbolts, M., Metabolic Engineering for microbial production of aromatic amino acids and derived compounds. *Metab Eng* **2001,** *3* (4), 289-300.

51. Rodriguez, A.; Martnez, J. A.; Flores, N.; Escalante, A.; Gosset, G.; Bolivar, F., Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Microb Cell Fact* **2014,** *13*, 126.

52. Vane, J.; Botting, R., The mechanism of action of aspirin. *Thromb Res* **2003,** *110* (5), 255-258.

53. Lin, Y.; Sun, X.; Yuan, Q.; Yan, Y., Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in *Escherichia coli*. *Metab Eng* **2014,** *23*, 62-69.

54. Noda, S.; Shirai, T.; Oyama, S.; Kondo, A., Metabolic design of a platform *Escherichia coli* strain producing various chorismate derivatives. *Metab Eng* **2016,** *33*, 119-129.

55. Nakagawa, A.; Minami, H.; Kim, J.-S.; Koyanagi, T.; Katayama, T.; Sato, F.; Kumagai, H., A bacterial platform for fermentative production of plant alkaloids. *Nat Commun* **2011,** *2*, 326.

56. Wu, J.; Du, G.; Zhou, J.; Chen, J., Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy. *Metab Eng* **2013,** *16*, 48-55.

57. Santos, C. N.; Koffas, M.; Stephanopoulos, G., Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab Eng* **2011,** *13* (4), 392-400.

58. Rodrigues, A. L.; Trachtmann, N.; Becker, J.; Lohanatha, A. F.; Blotenberg, J.; Bolten, C. J.; Korneli, C.; de Souza Lima, A. O.; Porto, L. M.; Sprenger, G. A.; Wittmann, C., Systems metabolic engineering of *Escherichia coli* for production of the antitumor drugs violacein and deoxyviolacein. *Metab Eng* **2013,** *20*, 29-41.

59. Rodrigues, A. L.; Becker, J.; de Souza Lima, A. O.; Porto, L. M.; Wittmann, C., Systems metabolic engineering of *Escherichia coli* for gram scale production of the antitumor drug deoxyviolacein from glycerol. *Biotechnol Bioeng* **2014,** *111* (11), 2280-2289.

60. Shumilin, I. A.; Kretsinger, R. H.; Bauerle, R. H., Crystal structure of phenylalanine-regulated 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*. *Structure* **1999,** *7* (7), 865-875.

61. Shumilin, I. A.; Zhao, C.; Bauerle, R.; Kretsinger, R. H., Allosteric inhibition of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase alters the coordination of both substrates. *J Mol Biol* **2002,** *320* (5), 1147-1156.

62. Hartmann, M.; Schneider, T. R.; Pfeil, A.; Heinrich, G.; Lipscomb, W. N.; Braus, G. H., Evolution of feedback-inhibited beta/alpha barrel isoenzymes by gene duplication and a single mutation. *Proc Natl Acad Sci U S A* **2003,** *100* (3), 862-867.

63. Cross, P. J.; Dobson, R. C.; Patchett, M. L.; Parker, E. J., Tyrosine latching of a regulatory gate affords allosteric control of aromatic amino acid biosynthesis. *J Biol Chem* **2011,** *286* (12), 10216-10224.

64. Webby, C. J.; Jiao, W.; Hutton, R. D.; Blackmore, N. J.; Baker, H. M.; Baker, E. N.; Jameson, G. B.; Parker, E. J., Synergistic allostery, a sophisticated regulatory network for the control of aromatic amino acid biosynthesis in *Mycobacterium tuberculosis*. *J Biol Chem* **2010,** *285* (40), 30567-30576.

65. Jensen, R. A.; Xie, G.; Calhoun, D. H.; Bonner, C. A., The correct phylogenetic relationship of KdsA (3-deoxy-D-*manno*-octulosonate 8-phosphate synthase) with one of two independently evolved classes of AroA (3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase). *J Mol Evol* **2002,** *54* (3), 416-423.

66. Subramaniam, P. S.; Xie, G.; Xia, T.; Jensen, R. A., Substrate ambiguity of 3-deoxy-D-*manno*-octulosonate 8-phosphate synthase from *Neisseria gonorrhoeae* in the context of its membership in a protein family containing a subset of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthases. *J Bacteriol* **1998,** *180* (1), 119-127.

67. Ahn, M.; Pietersma, A. L.; Schofield, L. R.; Parker, E. J., Mechanistic divergence of two closely related aldol-like enzyme-catalysed reactions. *Org Biomol Chem* **2005,** *3* (22), 4046-4049.

68. Wagner, T.; Shumilin, I. A.; Bauerle, R.; Kretsinger, R. H., Structure of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*: comparison of the Mn^{2+}∗2-phosphoglycolate and the Pb^{2+}∗2-phosphoenolpyruvate complexes and implications for catalysis. *J Mol Biol* **2000,** *301* (2), 389-399.

69. Chaudhury, S.; Abdulhameed, M. D. M.; Singh, N.; Tawa, G. J.; D'haeseleer, P. M.; Zemla, A. T.; Navid, A.; Zhou, C. E.; Franklin, M. C.; Cheung, J.; Rudolph, M. J.; Love, J.; Graf, J. F.; Rozak, D. A.; Dankmeyer, J. L.; Amemiya, K.; Daefler, S.; Wallqvist, A., Rapid countermeasure discovery against *Francisella tularensis* based on a metabolic network reconstruction. *PLOS ONE* **2013,** *8* (5), e63369.

70. Cross, P. J.; Pietersma, A. L.; Allison, T. M.; Wilson-Coutts, S. M.; Cochrane, F. C.; Parker, E. J., *Neisseria meningitidis* expresses a single 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase that is inhibited primarily by phenylalanine. *Protein Sci* **2013,** *22* (8), 1087-1099.

71. Schofield, L. R.; Patchett, M. L.; Parker, E. J., Expression, purification, and characterization of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase from *Pyrococcus furiosus*. *Protein Expr Purif* **2004,** *34* (1), 17-27.

72. Schofield, L. R.; Anderson, B. F.; Patchett, M. L.; Norris, G. E.; Jameson, G. B.; Parker, E. J., Substrate ambiguity and crystal structure of *Pyrococcus furiosus* 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase: an ancestral 3-deoxyald-2-ulosonate-phosphate synthase? *Biochemistry* **2005,** *44* (36), 11950-11962.

73. Zhou, L.; Wu, J.; Janakiraman, V.; Shumilin, I. A.; Bauerle, R.; Kretsinger, R. H., Structure and characterization of the 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase from *Aeropyrum pernix Bioorg Chem* **2012,** *40* (1), 79-86.

74. Wu, J.; Sheflyan, G. Y.; Woodard, R. W., *Bacillus subtilis* 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase revisited: resolution of two long-standing enigmas. *Biochem J* **2005,** *390* (Pt 2), 583.

75. Wu, J.; Woodard, R. W., New insights into the evolutionary links relating to the 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase subfamilies. *J Biol Chem* **2006,** *281* (7), 4042-4048.

76. Wu, J.; Howe, D. L.; Woodard, R. W., *Thermotoga maritima* 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP) synthase: the ancestral eubacterial DAHP synthase? *J Biol Chem* **2003,** *278* (30), 27525-27531.

77. Light, S. H.; Halavaty, A. S.; Minasov, G.; Shuvalova, L.; Anderson, W. F., Structural analysis of a 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase with an N-terminal chorismate mutase-like regulatory domain. *Protein Sci* **2012,** *21* (6), 887-895.

78. Nazmi, A. R.; Lang, E. J.; Bai, Y.; Allison, T. M.; Othman, M. H.; Panjikar, S.; Arcus, V. L.; Parker, E. J., Interdomain conformational changes provide allosteric regulation en route to chorismate. *J Biol Chem* **2016,** *291* (42), 21836-21847.

79. Burschowsky, D.; Thorbjørnsrud, H. V.; Heim, J. B.; Fahrig-Kamarauskaitė, J.; Würth-Roderer, K.; Kast, P.; Krengel, U., Inter-enzyme allosteric regulation of chorismate mutase in *Corynebacterium glutamicum*: Structural basis of feedback activation by Trp. *Biochemistry* **2017** (Web).

80. Light, S. H.; Anderson, W. F., The diversity of allosteric controls at the gateway to aromatic amino acid biosynthesis. *Protein Sci* **2013,** *22* (4), 395-404.

81. Hudson, G. S.; Rellos, P.; Davidson, B. E., Two promoters control the aroH gene of *Escherichia coli*. *Gene* **1991,** *102* (1), 87-91.

82. Krappmann, S.; Lipscomb, W. N.; Braus, G. H., Coevolution of transcriptional and allosteric regulation at the chorismate metabolic branch point of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **2000,** *97* (25), 13585-13590.

83. Cho, B. K.; Federowicz, S.; Park, Y. S.; Zengler, K.; Palsson, B. O., Deciphering the transcriptional regulatory logic of amino acid metabolism. *Nat Chem Biol* **2011,** *8* (1), 65-71.

84. König, V.; Pfeil, A.; Braus, G. H.; Schneider, T. R., Substrate and metal complexes of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Saccharomyces cerevisiae* provide new insights into the catalytic mechanism. *J Mol Biol* **2004,** *337* (3), 675-690.

85. Lang, E. J. M.; Heyes, L. C.; Jameson, G. B.; Parker, E. J., Calculated pKa variations expose dynamic allosteric communication networks. *J Am Chem Soc* **2016,** *138* (6), 2036-2045.

86. Webby, C. J.; Lott, J. S.; Baker, H. M.; Baker, E. N.; Parker, E. J., Crystallization and preliminary X-ray crystallographic analysis of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Mycobacterium tuberculosis*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2005,** *61* (Pt 4), 403-406.

87. Blackmore, N. J.; Reichau, S.; Jiao, W.; Hutton, R. D.; Baker, E. N.; Jameson, G. B.; Parker, E. J., Three sites and you are out: ternary synergistic allostery controls aromatic amino acid biosynthesis in *Mycobacterium tuberculosis*. *J Mol Biol* **2013,** *425* (9), 1582-1592.

88. Blackmore, N. J.; Nazmi, A. R.; Hutton, R. D.; Webby, M. N.; Baker, E. N.; Jameson, G. B.; Parker, E. J., Complex formation between two biosynthetic enzymes modifies the allosteric regulatory properties of both an example of molecular symbiosis. *J Biol Chem* **2015,** *290* (29), 18187-18198.

89. Munack, S.; Roderer, K.; Ökvist, M.; Kamarauskaite, J.; Sasso, S.; van Eerde, A.; Kast, P.; Krengel, U., Remote control by inter-enzyme allostery: a novel paradigm for regulation of the shikimate pathway. *J Mol Biol* **2016,** *428* (6), 1237-1255.

90. Siltberg-Liberles, J.; Martinez, A., Searching distant homologs of the regulatory ACT domain in phenylalanine hydroxylase. *Amino Acids* **2009,** *36* (2), 235-249.

91. Schuller, D. J.; Grant, G. A.; Banaszak, L. J., The allosteric ligand site in the Vmax-type cooperative enzyme phosphoglycerate dehydrogenase. *Nat Struct Mol Biol* **1995,** *2* (1), 69-76.

92. Aravind, L.; Koonin, E. V., Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J Mol Biol* **1999,** *287* (5), 1023-1040.

93. Cho, Y.; Sharma, V.; Sacchettini, J. C., Crystal structure of ATP phosphoribosyltransferase from *Mycobacterium tuberculosis*. *J Biol Chem* **2003,** *278* (10), 8333-8339.

94. Dumas, R.; Cobessi, D.; Robin, A. Y.; Ferrer, J. L.; Curien, G., The many faces of aspartate kinases. *Arch Biochem Biophys* **2012,** *519* (2), 186-193.

95. Fitzpatrick, P. F., Tetrahydropterin-dependent amino acid hydroxylases. *Annu Rev Biochem* **1999,** *68* (1), 355-381.

96. Jaffe, E. K.; Stith, L.; Lawrence, S. H.; Andrake, M.; Dunbrack, R. L., A new model for allosteric regulation of phenylalanine hydroxylase: implications for disease and therapeutics. *Arch Biochem Biophys* **2013,** *530* (2), 73-82.

97. Barak, Z.; Chipman, D. M., Allosteric regulation in acetohydroxyacid synthases (AHASs)--different structures and kinetic behavior in isozymes in the same organisms. *Arch Biochem Biophysics* **2012,** *519* (2), 167-174.

98. Lang, E. J.; Cross, P. J.; Mittelstädt, G.; Jameson, G. B.; Parker, E. J., Allosteric ACTion: the varied ACT domains regulating enzymes of amino-acid metabolism. *Curr Opin Struct Biol* **2014,** *29*, 102-111.

99. Devedjiev, Y.; Surendranath, Y.; Derewenda, U.; Gabrys, A.; Cooper, D. R.; Zhang, R.-g.; Lezondra, L.; Joachimiak, A.; Derewenda, Z. S., The structure and ligand binding properties of the *B. subtilis* ykoF gene product, a member of a novel family of thiamin/HMP-binding proteins. *J Mol Biol* **2004,** *343* (2), 395-406.

100. Cotton, R. G.; Gibson, F., The biosynthesis of phenylalanine and tyrosine; enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase. *Biochim et Biophys Acta* **1965,** *100* (1), 76-88.

101. Lee, A. Y.; Karplus, P. A.; Ganem, B.; Clardy, J., Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J Am Chem Soc* **1995,** *117* (12), 3627-3628.

102. Nester, E. W.; Lorence, J. H.; Nasser, D. S., An enzyme aggregate involved in the biosynthesis of aromatic amino acids in *Bacillus subtilis*. Its possible function in feedback regulation. *Biochemistry* **1967,** *6* (5), 1553-1563.

103. Davidson, B. E., Chorismate mutase-prephenate dehydratase from *Escherichia coli*. *Methods Enzymol* **1987,** *142*, 432-9.

104. Fischer, R. S.; Zhao, G.; Jensen, R. A., Cloning, sequencing, and expression of the P-protein gene (pheA) of *Pseudomonas stutzeri* in *Escherichia coli*: implications for

evolutionary relationships in phenylalanine biosynthesis. *J Gen Microbiol* **1991,** *137* (6), 1293-1301.

105. Xia, T.; Zhao, G.; Jensen, R. A., Loss of allosteric control but retention of the bifunctional catalytic competence of a fusion protein formed by excision of 260 base pairs from the 3' terminus of pheA from *Erwinia herbicola*. *Appl Environ Microbiol* **1992,** *58* (9), 2792-2798.

106. Gu, W.; Williams, D.; Aldrich, H.; Xie, G.; Gabriel, D.; Jensen, R., The AroQ and PheA domains of the bifunctional P-protein from *Xanthomonas campestris* in a context of genomic comparison. *Microb Comp Genomics* **1997,** *2*, 141 - 158.

107. Zhang, S.; Pohnert, G.; Kongsaeree, P.; Wilson, D. B.; Clardy, J.; Ganem, B., Chorismate mutase-prephenate dehydratase from *Escherichia coli*. Study of catalytic and regulatory domains using genetically engineered proteins. *J Biol Chem* **1998,** *273* (11), 6248-6253.

108. Turnbull, J.; Morrison, J. F.; Cleland, W. W., Kinetic studies on chorismate mutase-prephenate dehydrogenase from *Escherichia coli*: models for the feedback inhibition of prephenate dehydrogenase by L-tyrosine. *Biochemistry* **1991,** *30* (31), 7783-7788.

109. Xie, G.; Bonner, C. A.; Jensen, R. A., Cyclohexadienyl dehydrogenase from *Pseudomonas stutzeri* exemplifies a widespread type of tyrosine-pathway dehydrogenase in the TyrA protein family. *Comp Biochem Physiol C Toxicol Pharmacol* **2000,** *125* (1), 65-83.

110. Lutke-Eversloh, T.; Stephanopoulos, G., Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of *Escherichia coli*: generation and characterization of tyrosine-insensitive mutants. *Appl Environ Microbiol* **2005,** *71* (11), 7224-7228.

111. Sun, W.; Shahinas, D.; Bonvin, J.; Hou, W.; Kimber, M. S.; Turnbull, J.; Christendat, D., The crystal structure of *Aquifex aeolicus* prephenate dehydrogenase reveals the mode of tyrosine inhibition. *J Biol Chem* **2009,** *284* (19), 13223-13232.

112. Champney, W. S.; Jensen, R. A., The enzymology of prephenate dehydrogenase in *Bacillus subtilis*. *J Biol Chem* **1970,** *245* (15), 3763-3770.

113. Chiu, H. J.; Abdubek, P.; Astakhova, T.; Axelrod, H. L.; Carlton, D.; Clayton, T.; Das, D.; Deller, M. C.; Duan, L.; Feuerhelm, J.; Grant, J. C.; Grzechnik, A.; Han, G. W.; Jaroszewski, L.; Jin, K. K.; Klock, H. E.; Knuth, M. W.; Kozbial, P.; Krishna, S. S.; Kumar, A.; Marciano, D.; McMullan, D.; Miller, M. D.; Morse, A. T.; Nigoghossian, E.; Okach, L.; Reyes, R.; Tien, H. J.; Trame, C. B.; van den Bedem, H.; Weekes, D.; Xu, Q.; Hodgson, K. O.; Wooley, J.; Elsliger, M. A.; Deacon, A. M.; Godzik, A.; Lesley, S. A.; Wilson, I. A., The structure of *Haemophilus influenzae* prephenate

dehydrogenase suggests unique features of bifunctional TyrA enzymes. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2010,** *66* (Pt 10), 1317-1325.

114. Lee, A.; Stewart, J. D.; Clardy, J.; Ganem, B., New insight into the catalytic mechanism of chorismate mutases from structural studies. *Chem Biol* **1995,** *2* (4), 195-203.

115. MacBeath, G.; Kast, P.; Hilvert, D., A small, thermostable, and monofunctional chorismate mutase from the archeon *Methanococcus jannaschii*†. *Biochemistry* **1998,** *37* (28), 10062-10073.

116. Chook, Y. M.; Ke, H.; Lipscomb, W. N., Crystal structures of the monofunctional chorismate mutase from *Bacillus subtilis* and its complex with a transition state analog. *Proc Nat Acad Sci* **1993,** *90* (18), 8600-8603.

117. Cross, P. J.; Allison, T. M.; Dobson, R. C. J.; Jameson, G. B.; Parker, E. J., Engineering allosteric control to an unregulated enzyme by transfer of a regulatory domain. *Proc Natl Acad Sci U S A* **2013,** *110* (6), 2111-2116.

118. Shumilin, I. A.; Bauerle, R.; Wu, J.; Woodard, R. W.; Kretsinger, R. H., Crystal structure of the reaction complex of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Thermotoga maritima* refines the catalytic mechanism and indicates a new mechanism of allosteric regulation. *J Mol Biol* **2004,** *341* (2), 455-466.

119. Cross, P. J.; Parker, E. J., Allosteric inhibitor specificity of *Thermotoga maritima* 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase. *FEBS Lett.* **2013,** *587* (18), 3063-3068.

120. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res* **2004,** *14* (6), 1188-90.

121. Rambo, R. P.; Tainer, J. A., Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr Opin Struct Biol* **2010,** *20* (1), 128-137.

122. Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A., X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **2007,** *40* (03), 191-285.

123. Kabsch, W., XDS. *Acta Crystallogr D* **2010,** *66* (2), 125-132.

124. Kabsch, W., Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D* **2010,** *66* (2), 133-144.

125. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallographic D* **2011,** *67* (Pt 4), 235-242.

126. Vagin, A.; Teplyakov, A., MOLREP: an automated program for molecular replacement. *Appl Crystallogr* **1997,** *30* (6), 1022-1025.

127. Henry, E. A.; Devereux, R.; Maki, J. S.; Gilmour, C. C.; Woese, C. R.; Mandelco, L.; Schauder, R.; Remsen, C. C.; Mitchell, R., Characterization of a new thermophilic sulfate-reducing bacterium *Thermodesulfovibrio yellowstonii*, gen. nov. and sp. nov.: its phylogenetic relationship to Thermodesulfobacterium commune and their origins deep within the bacterial domain. *Arch Microbiol* **1994,** *161* (1), 62-69.

128. Cole, J. L.; Lary, J. W.; T, P. M.; Laue, T. M., Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol* **2008,** *84*, 143-179.

129. Schuck, P., Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **2000,** *78* (3), 1606-1619.

130. Lebowitz, J.; Lewis, M. S.; Schuck, P., Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Sci* **2002,** *11* (9), 2067-2079.

131. Grant, G. A., The ACT domain: a small molecule binding domain and its role as a common regulatory element. *J Biol Chem* **2006,** *281* (45), 33825-33829.

132. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **2007,** *372* (3), 774-797.

133. Perica, T.; Marsh, Joseph A.; Sousa, Filipa L.; Natan, E.; Colwell, Lucy J.; Ahnert, Sebastian E.; Teichmann, Sarah A., The emergence of protein complexes: quaternary structure, dynamics and allostery. *Biochem Soc Trans* **2012,** *40* (3), 475-491.

134. Nazmi, A. R.; Schofield, L. R.; Dobson, R. C. J.; Jameson, G. B.; Parker, E. J., Destabilization of the homotetrameric assembly of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from the hyperthermophile *Pyrococcus furiosus* enhances enzymatic activity. *J Mol Biol* **2014,** *426* (3), 656-673.

135. Cansu, S.; Doruker, P., Dimerization affects collective dynamics of triosephosphate isomerase. *Biochemistry* **2008,** *47* (5), 1358-1368.

136. Park, C.; Raines, R. T., Dimer formation by a "monomeric" protein. *Protein Sci* **2000,** *9* (10), 2026-2033.

137. Pereira-Leal, J. B.; Levy, E. D.; Kamp, C.; Teichmann, S. A., Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol* **2007,** *8* (4), R51.

138. Devenish, S. R. A.; Gerrard, J. A., The role of quaternary structure in (beta/alpha)8-barrel proteins: evolutionary happenstance or a higher level of structure-function relationships? *Org Biomol Chem* **2009,** *7* (5), 833-839.

139. Cross, P. J.; Heyes, L. C.; Zhang, S.; Nazmi, A. R.; Parker, E. J., The functional unit of *Neisseria meningitidis* 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase is dimeric. *PLOS ONE* **2016,** *11* (2), e0145187.

140. Jiao, W.; Blackmore, N. J.; Nazmi, A. R.; Parker, E. J., Quaternary structure is an essential component that contributes to the sophisticated allosteric regulation mechanism in a key enzyme from *Mycobacterium tuberculosis*. *PLOS ONE* **2017,** *12* (6), e0180052.

141. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature* **2014,** *508* (7496), 331-339.

142. Goodey, N. M.; Benkovic, S. J., Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* **2008,** *4* (8), 474-482.

143. Swain, J. F.; Gierasch, L. M., The changing landscape of protein allostery. *Curr Opin Struct Biol* **2006,** *16* (1), 102-108.

144. Lee, J.; Natarajan, M.; Nashine, V. C.; Socolich, M.; Vo, T.; Russ, W. P.; Benkovic, S. J.; Ranganathan, R., Surface sites for engineering allosteric control in proteins. *Science* **2008,** *322* (5900), 438-442.

145. Ma, B.; Tsai, C. J.; Haliloglu, T.; Nussinov, R., Dynamic allostery: linkers are not merely flexible. *Structure* **2011,** *19* (7), 907-917.

146. Doolittle, R. F., The multiplicity of domains in proteins. *Annu Rev Biochem* **1995,** *64*, 287-314.

147. Patthy, L., Modular assembly of genes and the evolution of new functions. *Genetica* **2003,** *118* (2-3), 217-231.

148. Vogel, C.; Bashton, M.; Kerrison, N. D.; Chothia, C.; Teichmann, S. A., Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* **2004,** *14* (2), 208-216.

149. Chothia, C., Proteins. One thousand families for the molecular biologist. *Nature* **1992,** *357* (6379), 543-544.

150. Peracchi, A.; Mozzarelli, A., Exploring and exploiting allostery: models, evolution, and drug targeting. *Biochim Biophys Acta* **2011,** *1814* (8), 922-933.

151. Bashton, M.; Chothia, C., The generation of new protein functions by the combination of domains. *Structure* **2007,** *15* (1), 85-99.

152. Ostermeier, M.; Benkovic, S. J., Evolution of protein function by domain swapping. *Adv Protein Chem* **2000,** *55*, 29-77.

153. Pasek, S.; Risler, J. L.; Brezellec, P., Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* **2006,** *22* (12), 1418-1423.

154. Pawson, T.; Nash, P., Assembly of cell regulatory systems through protein interaction domains. *Science* **2003,** *300* (5618), 445-452.

155. Hatley, M. E.; Lockless, S. W.; Gibson, S. K.; Gilman, A. G.; Ranganathan, R., Allosteric determinants in guanine nucleotide-binding proteins. *Proc Natl Acad Sci U S A* **2003,** *100* (24), 14445-14450.

156. Ostermeier, M., Engineering allosteric protein switches by domain insertion. *Protein Eng Des Sel* **2005,** *18* (8), 359-364.

157. Betton, J.-M.; Jacob, J. P.; Hofnung, M.; Broome-Smith, J. K., Creating a bifunctional protein by insertion of b-lactamase into the maltodextrin-binding protein. *Nat Biotechnol* **1997,** *15* (12), 1276-1279.

158. Guntas, G.; Ostermeier, M., Creation of an allosteric enzyme by domain insertion. *J Mol Biol* **2004,** *336* (1), 263-273.

159. Dueber, J. E.; Mirsky, E. A.; Lim, W. A., Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotech* **2007,** *25* (6), 660-662.

160. Dueber, J. E.; Yeh, B. J.; Chak, K.; Lim, W. A., Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **2003,** *301* (5641), 1904-1908.

161. Allison, T. M.; Hutton, R. D.; Jiao, W.; Gloyne, B. J.; Nimmo, E. B.; Jameson, G. B.; Parker, E. J., An extended β7α7 substrate-binding loop is essential for efficient catalysis by 3-deoxy-D-*manno*-octulosonate 8-phosphate synthase. *Biochemistry* **2011,** *50* (43), 9318-9327.

162. Schwartz, T. W.; Holst, B., Allosteric enhancers, allosteric agonists and ago-allosteric modulators: where do they bind and how do they act? *Trends Pharmacol Sci* **2007,** *28* (8), 366-373.

163. Nussinov, R.; Tsai, C. J., Allostery in disease and in drug discovery. *Cell* **2013,** *153* (2), 293-305.

164. Zorn, J. A.; Wells, J. A., Turning enzymes ON with small molecules. *Nat Chem Biol* **2010,** *6* (3), 179-188.

165. Conn, P. J.; Christopoulos, A.; Lindsley, C. W., Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nat Rev Drug Discov* **2009,** *8* (1), 41-54.

166. Wood, M. R.; Hopkins, C. R.; Brogan, J. T.; Conn, P. J.; Lindsley, C. W., "Molecular switches" on mglur allosteric ligands that modulate modes of pharmacology. *Biochemistry* **2011,** *50* (13), 2403-2410.

167. Nussinov, R.; Tsai, C. J.; Csermely, P., Allo-network drugs: harnessing allostery in cellular networks. *Trends Pharmacol Sci* **2011,** *32* (12), 686-693.

168. Sadowsky, J. D.; Burlingame, M. A.; Wolan, D. W.; McClendon, C. L.; Jacobson, M. P.; Wells, J. A., Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc Nat Acad Sci* **2011,** *108* (15), 6056-6061.

169. Naumann, M.; Sokolova, O.; Tegtmeyer, N.; Backert, S., *Helicobacter pylori*: a paradigm pathogen for subverting host cell signal transmission. *Trends Microbiol* **2017,** *25* (4), 316-328.

170. Salama, N. R.; Hartung, M. L.; Muller, A., Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*. *Nat Rev Micro* **2013,** *11* (6), 385-399.

171. Amieva, M.; Peek, R. M., Jr., Pathobiology of *Helicobacter pylori*-induced gastric cancer. *Gastroenterology* **2016,** *150* (1), 64-78.

172. Graham, D. Y.; Fischbach, L., *Helicobacter pylori* treatment in the era of increasing antibiotic resistance. *Gut* **2010**, *59* (8), 1143-1153.

173. Megraud, F., *H pylori* antibiotic resistance: prevalence, importance, and advances in testing. *Gut* **2004,** *53* (9), 1374-1384.

174. Fersht, A., *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. W. H. Freeman: **1999.**

175. Li, P. P.; Li, D. F.; Liu, D.; Liu, Y. M.; Liu, C.; Liu, S. J., Interaction between DAHP synthase and chorismate mutase endows new regulation on DAHP synthase activity in Corynebacterium glutamicum. *Appl Microbiol Biotechnol* **2013**, 1-8.

176. Webby, C. J., Structural & functional characterization of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase from *Helicobacter Pylori* & *Mycobacterium Tuberculosis.* Massey University, Turitea: **2006**.

177. Webby, C. J.; Patchett, M. L.; Parker, E. J., Characterization of a recombinant type II 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Helicobacter pylori*. *Biochem J* **2005,** *390* (1), 223-230.

178. Webby, C. J.; Baker, H. M.; Lott, J. S.; Baker, E. N.; Parker, E. J., The structure of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase from *Mycobacterium tuberculosis* reveals a common catalytic scaffold and ancestry for type i and type ii enzymes. *J Mol Biol* **2005,** *354* (4), 927-939.

179. Gosset, G.; Bonner, C. A.; Jensen, R. A., Microbial origin of plant-type 2 keto-3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthases, exemplified by the chorismate- and tryptophan-regulated enzyme from *Xanthomonas campestris*. *J Bacteriol* **2001,** *183*(13), 4061-4070.

180. Yoo, J. C.; Hee-Jae, B.; Eun-Ha, L.; Sung-Jun, K.; Jung-Jun, L., Purification and characteristics of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthetase from *Streptomyces caespitosus*. *Kor J Microbiol* **1993,** *31* (4), 340-345.

181. Walker, G. E.; Dunbar, B.; Hunter, I. S.; Nimmo, H. G.; Coggins, J. R., Evidence for a novel class of microbial 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase in *Streptomyces coelicolor*, *Streptomyces rimosus* and *Neurospora crassa*. *Microbiology* **1996,** *142 (Pt 8)*, 1973-1982.

182. Stuart, F.; Hunter, I. S., Purification and characterization of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Streptomyces rimosus*. *Biochim Biophys Acta* **1993,** *1161* (2-3), 209-215.

183. Doy, C. H., The kinetics of a purified form of 3-deoxy-D-*arabino* heptulosonate-7-phosphate synthase (tryptophan) from *Neurospora crassa*. *FEBS J* **1979,** *98* (2), 431-440.

184. Murphy, M. F.; Katz, E., Regulatory control of 3-deoxy-D-*arabino*-heptulosonic acid 7-phosphate synthetase in *Streptomyces antibioticus*. *Can J Microbiol***1980,** *26* (8), 874-880.

185. Fischer, H.; de Oliveira Neto, M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F., Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J Appl Crystallogr* **2010,** *43* (1), 101-109.

186. Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I., ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **2017,** *50* (Pt 4), 1212-1225.

187. Svergun, D.; Barberato, C.; Koch, M. H., CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **1995,** *28* (6), 768-773.

188. Jiao, W.; Hutton, R. D.; Cross, P. J.; Jameson, G. B.; Parker, E. J., Dynamic cross-talk among remote binding sites: the molecular basis for unusual synergistic allostery. *J Mol Biol* **2012,** *415* (4), 716-26.

189. Zhang, S.; Kongsaeree, P.; Clardy, J.; Wilson, D. B.; Ganem, B., Site-directed mutagenesis of monofunctional chorismate mutase engineered from the *E. coli* P-protein. *Bioorg Med Chem* **1996,** *4* (7), 1015-1020.

190. Calhoun, D.; Bonner, C.; Gu, W.; Xie, G.; Jensen, R., The emerging periplasm-localized subclass of AroQ chorismate mutases, exemplified by those from Salmonella typhimurium and *Pseudomonas aeruginosa*. *Genome Biol* **2001,** *2* (8), research0030.1 - research0030.16.

191. Xia, T.; Song, J.; Zhao, G.; Aldrich, H.; Jensen, R., The aroQ encoded monofunctional chorismate mutase (CM-F) protein is a periplasmic enzyme in *Erwinia herbicola*. *J Bacteriol* **1993,** *175*, 4729 - 4737.

192. Xia, T.; Zhao, G.; Jensen, R. A., The pheA/tyrA/aroF region from *Erwinia herbicola*: An emerging comparative basis for analysis of gene organization and regulation in enteric bacteria. *J Mol Evol* **1993,** *36* (2), 107-120.

193. Fazel, A. M.; Bowen, J. R.; Jensen, R. A., Arogenate (pretyrosine) is an obligatory intermediate of L-tyrosine biosynthesis: confirmation in a microbial mutant. *Proc Nat Acad Sci* **1980,** *77* (3), 1270-1273.

194. Xue, Y.; Lipscomb, W. N., Location of the active site of allosteric chorismate mutase from *Saccharomyces cerevisiae*, and comments on the catalytic and regulatory mechanisms. *Proc Nat Acad Sci* **1995,** *92* (23), 10595-10598.

195. Sasso, S.; Ramakrishnan, C.; Gamper, M.; Hilvert, D.; Kast, P., Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*. *FEBS J* **2005,** *272* (2), 375-389.

196. Jones, D. T., Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **1999,** *292* (2), 195-202.

197. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **2006,** *1* (6), 2876-2890.

198. Andrade, M.; Chacon, P.; Merelo, J.; Moran, F., Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng Des Sel* **1993,** *6* (4), 383-390.

199. McPhillips, T. M.; McPhillips, S. E.; Chiu, H.-J.; Cohen, A. E.; Deacon, A. M.; Ellis, P. J.; Garman, E.; Gonzalez, A.; Sauter, N. K.; Phizackerley, R. P.; Soltis, S. M.; Kuhn, P., Blu-Ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J Synchrotron Radiat* **2002,** *9* (6), 401-406.

200. Karplus, P. A.; Diederichs, K., Linking Crystallographic model and data quality. *Science* **2012,** *336* (6084), 1030-1033.

201. Stein, N., CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* **2008,** *41* (3), 641-643.

202. Terwilliger, T., Maximum-likelihood density modification. *Acta Crystallogr D* **2000,** *56* (8), 965-972.

203. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **1997,** *53* (3), 240-255.

204. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **2004,** *60* (Pt 12 Pt 1), 2126-32

205. Zaitseva, J.; Lu, J.; Olechoski, K. L.; Lamb, A. L., Two crystal structures of the isochorismate pyruvate lyase from *Pseudomonas aeruginosa*. *J Biol Chem* **2006,** *281* (44), 33441-33449.

206. Olucha, J.; Ouellette, A. N.; Luo, Q.; Lamb, A. L., pH Dependence of catalysis by *Pseudomonas aeruginosa* isochorismate–pyruvate lyase: implications for transition state stabilization and the role of lysine 42. *Biochemistry* **2011,** *50* (33), 7198-7207.

207. Blackmore, N. J. The regulation of 3-deoxy-D-*arabino*-heptulosonate 7 phosphate synthase from *Mycobacterium tuberculosis*. University of Canterbury, **2015**.

208. Lamb, A. L., Pericyclic reactions catalyzed by chorismate-utilizing enzymes. *Biochemistry* **2011,** *50* (35), 7476-7483.

209. Weng, J. K.; Noel, J. P., The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb Symp Quant Biol* **2012,** *77*, 309-320.

210. Ferrer, S.; Marti, S.; Moliner, V.; Tunon, I.; Bertran, J., Understanding the different activities of highly promiscuous MbtI by computational methods. *Phys Chem Chem Phys* **2012,** *14* (10), 3482-3489.

211. Rost, B.; Yachdav, G.; Liu, J., The predictprotein server. *Nucleic Acids Res* **2004,** *32* (suppl_2), W321-W326.

212. Smith, N.; Roitberg, A. E.; Rivera, E.; Howard, A.; Holden, M. J.; Mayhew, M.; Kaistha, S.; Gallagher, D. T., Structural analysis of ligand binding and catalysis in chorismate lyase. *Arch Biochem Biophys* **2006,** *445* (1), 72-80.

213. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007,** *450* (7172), 964-972.

214. Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y., Creating new fluorescent probes for cell biology. *Nat Rev Mol Cell Biol* **2002,** *3* (12), 906-918.

215. Chen, Y.; Tsao, K.; Keillor, J. W., Fluorogenic protein labelling: a review of photophysical quench mechanisms and principles of fluorogen design. *Can J Chem* **2014,** *93* (4), 389-398.

216. Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y., The fluorescent toolbox for assessing protein location and function. *Science* **2006,** *312* (5771), 217-224.

217. Jares-Erijman, E. A.; Jovin, T. M., FRET imaging. *Nat Biotech* **2003,** *21* (11), 1387-1395.

218. Rashidian, M.; Dozier, J. K.; Distefano, M. D., Enzymatic labeling of proteins: techniques and approaches. *Bioconjug Chem* **2013,** *24* (8), 1277-1294.

219. Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S., Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci U S A* **2000,** *97* (10), 5179-5184.

220. Tsien, R. Y., The green fluorescent protein. *Annu Rev Biochem* **1998,** *67* (1), 509-544.

221. Keppler, A.; Pick, H.; Arrivoli, C.; Vogel, H.; Johnsson, K., Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc Natl Acad Sci U S A* **2004,** *101* (27), 9955-9959.

222. Miller, L. W.; Cai, Y.; Sheetz, M. P.; Cornish, V. W., In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nat Meth* **2005,** *2* (4), 255-257.

223. Oteng-Pabi, S. K.; Pardin, C.; Stoica, M.; Keillor, J. W., Site-specific protein labelling and immobilization mediated by microbial transglutaminase. *Chem Commun* **2014,** *50* (50), 6604-6606.

224. Duckworth, B. P.; Zhang, Z.; Hosokawa, A.; Distefano, M. D., Selective labeling of proteins by using protein farnesyltransferase. *ChemBioChem* **2007,** *8* (1), 98-105.

225. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998,** *281* (5374), 269-272.

226. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y., New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. *J Am Chem Soc* **2002,** *124* (21), 6063-6076.

227. Langhorst, M. F.; Genisyuerek, S.; Stuermer, C. A. O., Accumulation of FlAsH/Lumio Green in active mitochondria can be reversed by β-mercaptoethanol for specific staining of tetracysteine-tagged proteins. *Histochem Cell Biol* **2006,** *125* (6), 743.

228. Nikic, I.; Plass, T.; Schraidt, O.; Szymanski, J.; Briggs, J. A.; Schultz, C.; Lemke, E. A., Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew Chem* **2014,** *53* (8), 2245-2249.

229. Kim, J.; Seo, M. H.; Lee, S.; Cho, K.; Yang, A.; Woo, K.; Kim, H. S.; Park, H. S., Simple and efficient strategy for site-specific dual labeling of proteins for single-molecule fluorescence resonance energy transfer analysis. *Anal Chem* **2013,** *85* (3), 1468-1474.

230. Margittai, M.; Widengren, J.; Schweinberger, E.; Schröder, G. F.; Felekyan, S.; Haustein, E.; König, M.; Fasshauer, D.; Grubmüller, H.; Jahn, R.; Seidel, C. A. M., Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc Nat Acad Sci* **2003,** *100* (26), 15516-15521.

231. Kuzmenkina, E. V.; Heyes, C. D.; Nienhaus, G. U., Single-molecule FRET study of denaturant induced unfolding of RNase H. *J Mol Biol* **2006,** *357* (1), 313-324.

232. Ohtsuka, T.; Ota, M.; Nio, N.; Motoki, M., Comparison of substrate specificities of transglutaminases using synthetic peptides as acyl donors. *Biosci Biotechnol Biochem* **2000,** *64* (12), 2608-2613.

233. Griffin, M.; Casadio, R.; Bergamini, C. M., Transglutaminases: nature's biological glues. *Biochem J* **2002,** *368* (Pt 2), 377-396.

234. Takahashi, N.; Takahashi, Y.; Putnam, F. W., Primary structure of blood coagulation factor XIIIa (fibrinoligase, transglutaminase) from human placenta. *Proc Nat Acad Sci* **1986,** *83* (21), 8019-8023.

235. Chhabra, A.; Verma, A.; Mehta, K., Tissue transglutaminase promotes or suppresses tumors depending on cell context. *Anticancer Res* **2009,** *29* (6), 1909-1919.

236. Oteng-Pabi, S. K.; Keillor, J. W., Continuous enzyme-coupled assay for microbial transglutaminase activity. *Anal Biochem* **2013,** *441* (2), 169-173.

237. Sugimura, Y.; Yokoyama, K.; Nio, N.; Maki, M.; Hitomi, K., Identification of preferred substrate sequences of microbial transglutaminase from *Streptomyces mobaraensis* using a phage-displayed peptide library. *Arch Biochem Biophys* **2008,** *477* (2), 379-383.

238. Taki, M.; Shiota, M.; Taira, K., Transglutaminase-mediated N- and C-terminal fluorescein labeling of a protein can support the native activity of the modified protein. *Protein Eng Des Sel* **2004,** *17* (2), 119-126.

239. Taraska, J. W.; Zagotta, W. N., Fluorescence applications in molecular neurobiology. *Neuron* **2010,** *66* (2), 170-189.

240. Haugland, R. P., *The handbook: a guide to fluorescent probes and labeling technologies*. Molecular probes: **2005**.

241. Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **2005,** *41* (1), 207-234.

242. Folk, J.; Cole, P., Transglutaminase: mechanistic features of the active site as determined by kinetic and inhibitor studies. *Biochim Biophys Acta* **1966,** *122* (2), 244-264.

243. Reginsson, G. W.; Schiemann, O., Studying biomolecular complexes with pulsed electron-electron double resonance spectroscopy. *Biochem Soc Trans* **2011,** *39* (1), 128-139.

244. Schiemann, O.; Prisner, T. F., Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Q Rev Biophys* **2007,** *40* (1), 1-53.

245. Hagelueken, G.; Ingledew, W. J.; Huang, H.; Petrovic-Stojanovska, B.; Whitfield, C.; ElMkami, H.; Schiemann, O.; Naismith, J. H., PELDOR spectroscopy distance fingerprinting of the octameric outer-membrane protein Wza from *Escherichia coli*. *Angew Chem* **2009,** *48* (16), 2904-2906.

246. Schiemann, O.; Piton, N.; Mu, Y.; Stock, G.; Engels, J. W.; Prisner, T. F., A PELDOR-based nanometer distance ruler for oligonucleotides. *J Am Chem Soc* **2004,** *126* (18), 5722-5729.

247. Jeschke, G., DEER Distance Measurements on Proteins. *Annu Rev Phys Chem* **2012,** *63* (1), 419-446.

248. Hubbell, W. L.; Cafiso, D. S.; Altenbach, C., Identifying conformational changes with site-directed spin labeling. *Nat Struct Biol* **2000,** *7* (9), 735-739.

249. Borbat, P. P.; McHaourab, H. S.; Freed, J. H., Protein structure determination using long-distance constraints from double-quantum coherence ESR: Study of T4 lysozyme. *J Am Chem Soc* **2002,** *124* (19), 5304-5314.

250. Jeschke, G.; Polyhach, Y.; Bordignon, E., Multiscale Modeling of Macromolecules–MMM. *ETH Zurich* **2010**.

251. Polyhach, Y.; Bordignon, E.; Jeschke, G., Rotamer libraries of spin labelled cysteines for protein studies. *Phys Chem Chem Phys* **2011,** *13* (6), 2356-2366.

252. Jeschke, G.; Chechik, V.; Ionita, P.; Godt, A.; Zimmermann, H.; Banham, J.; Timmel, C. R.; Hilger, D.; Jung, H., DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl Magn Reson* **2006,** *30* (3), 473-498.

253. Guy, J.; Castonguay, R.; Campos-Reales Pineda, N. B.; Jacquier, V.; Caron, K.; Michnick, S. W.; Keillor, J. W., De novo helical peptides as target sequences for a specific, fluorogenic protein labelling strategy. *Mol Biosyst* **2010,** *6* (6), 976-987.

254. Guy, J.; Caron, K.; Dufresne, S.; Michnick, S. W.; Skene, W. G.; Keillor, J. W., Convergent preparation and photophysical characterization of dimaleimide dansyl fluorogens: elucidation of the maleimide fluorescence quenching mechanism. *J Am Chem Soc* **2007,** *129* (39), 11969-11977.

255. Lang, E. J. M. Insights into enzyme allosteric inhibition mechanisms using computational studies. University of Canterbury, **2016**.

256. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **2011,** *7*, 539.

257. Robert, X.; Gouet, P., Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **2014,** *42* (Web Server issue), W320-4.

258. Finn, R. D.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; Salazar, G. A.; Tate, J.; Bateman, A., The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **2016,** *44* (D1), D279-D285.

259. Gouy, M.; Guindon, S.; Gascuel, O., SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **2010,** *27* (2), 221-224.

260. Schrödinger, L., The PyMOL molecular graphics system, version 1.8 Schrödinger, LLC. 2016.

261. Landy, A., Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* **1989,** *58*, 913-949.

262. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. e.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., *Protein identification and analysis tools on the ExPASy server*. Springer: **2005**.

263. Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **1976,** *72* (1–2), 248-254.

264. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I., PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **2003,** *36* (5), 1277-1282.

265. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr D* **2010,** *66* (Pt 4), 486-501.

266. Mandal, A.; Hilvert, D., Charge optimization increases the potency and selectivity of a chorismate mutase inhibitor. *J Am Chem Soc* **2003,** *125* (19), 5598-5599.

267. Andrews, P.; Smith, G. D.; Young, I., Transition-state stabilization and enzymic catalysis. Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate. *Biochemistry* **1973,** *12* (18), 3492-3498.

268. Zhang, X.; Zhang, X.; Bruice, T. C., A definitive mechanism for chorismate mutase. *Biochemistry* **2005,** *44* (31), 10443-8.

269. Guilford, W. J.; Copley, S. D.; Knowles, J. R., The mechanism of the chorismate mutase reaction. *J Am Chem Soc* **1987,** *109* (16), 5013-5019.

270. Copley, S. D.; Knowles, J. R., The conformational equilibrium of chorismate in solution: implications for the mechanism of the non-enzymic and the enzyme-catalyzed rearrangement of chorismate to prephenate. *J Am Chem Soc* **1987,** *109* (16), 5008-5013.

271. Kienhöfer, A.; Kast, P.; Hilvert, D., Selective stabilization of the chorismate mutase transition state by a positively charged hydrogen bond donor. *J Am Chem Soc* **2003,** *125* (11), 3206-3207.

272. Hur, S.; Bruice, T. C., The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc Natl Acad Sci* **2003,** *100* (21), 12015-12020.

273. Hur, S.; Bruice, T. C., Comparison of formation of reactive conformers (NACs) for the Claisen rearrangement of chorismate to prephenate in water and in the E. coli mutase: the efficiency of the enzyme catalysis. *J Am Chem Soc* **2003,** *125* (19), 5964-72.

274. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** *23* (21), 2947-8.